Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity*

Swati Dhingra
CEP, London School of Economics & CEPR

John Morrow
Essex and CEP, London School of Economics

This Draft: Oct 31, 2016

Abstract

Empirical work has drawn attention to the high degree of productivity differences within industries, and its role in resource allocation. This paper examines the allocational efficiency of such markets. Productivity differences introduce two new margins of potential inefficiency: selection of the right distribution of firms and allocation of the right quantities across firms. We show that these considerations impact welfare and policy analysis. Market power across firms leads to distortions in resource allocation. Demand-side elasticities determine how resources are misallocated and when increased competition from market expansion provides welfare gains.

JEL Codes: F1, L1, D6.
Keywords: Efficiency, Productivity, Social welfare, Demand elasticity, Markups.

Acknowledgments. We thank Bob Staiger for continued guidance and Steve Redding for encouragement. We are grateful to G Alessandria, C Arkolakis, R Armenter, A Bernard, S Chatterjee, D Chor, S Durlauf, C Engel, T Fally, R Feenstra, K Head, W Keller, J Lin, E Ornelas, G Ottaviano, M Parenti, N Pavcnik, T Sampson, D Sturm, J Thisse, J Van Reenen, A Weinberger, B Zissimos and M Zhu for insightful comments, K Russ and A Rodriguez-Clare for AEA discussions and T Besley for advice. This paper has benefited from helpful comments of participants at AEA 2011 and 2013, CEPR-ERWIT, CEPR-IO, Columbia, Davis, DIME-ISGEP 2010, ETSG 2012, Georgetown, Harvard KSG, HSE St Petersburg, ISI, FIW, LSE, Louvain, Mannheim, Maryland, NBER, Oxford, Philadelphia Fed, Princeton, Toronto, Virginia Darden, Wisconsin and Yale. Swati thanks the IES (Princeton) for their hospitality. We acknowledge the financial support from Portuguese national funds by FCT (Fundacao para a Ciencia e a Tecnologia) project PTDC/EGE-ECO/122115/2010. A preliminary draft was a dissertation chapter at Wisconsin in 2010.

*The first line is the title of Dixit and Stiglitz (1977). Contact: s.dhingra@lse.ac.uk; j.morrow1@lse.ac.uk.

1 Introduction

Empirical work has drawn attention to the high degree of heterogeneity in firm productivity, and the constant reallocation of resources across different firms.[1] The focus on productivity differences has provided new insights into market outcomes such as industrial productivity, firm pricing and welfare gains from policy changes.[2] When firms differ in productivity, the distribution of resources across firms also affects the allocational efficiency of markets. In a recent survey, Syverson (2011) notes the gap between social benefits and costs across firms has not been adequately examined, and this limited understanding has made it difficult to implement policies to reduce distortions (pp. 359). This paper examines allocational efficiency in markets where firms differ in productivity. 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?

Symmetric firm models explain when resource allocation is efficient by examining the trade-off between quantity and product variety in imperfectly competitive markets.[3] When firms differ in productivity, we must also ask which types of firms should produce and which should be shut down. Firm differences in productivity introduce two new margins of potential inefficiency: selection of the right distribution of firms and allocation of the right quantities across firms. For example, it could be welfare-improving to skew resources towards firms with lower costs (to obtain more output for a given level of resources) or towards firms with higher costs to preserve variety (because there is a real resource cost to introduce varieties and this cost is sunk once a variety is introduced). Furthermore, differences in market power across firms lead to new trade-offs between variety and quantity. These considerations impact optimal policy rules in a fundamental way, distinct from markets with symmetric costs. One contribution of the paper is to understand how these considerations affect welfare and policy analysis.

A second contribution of the paper is to show when increased competition improves welfare and efficiency. When market allocations are inefficient, increased competition (from trade or growth) may exacerbate distortions and lead to welfare losses (Helpman and Krugman 1985). A second-best world offers no guarantee of welfare gains from trade. But, by creating larger, more competitive markets, trade may reduce the distortions associated with imperfect competition and provide welfare gains (Krugman 1987). This insight is even more relevant in a heterogeneous cost environment because of new sources of potential inefficiency. We explain

[1] Example, Bartelsman and Doms (2000); Tybout (2003); Feenstra (2006); Bernard, Jensen, Redding and Schott (2007).

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

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

when integration provides welfare gains by aligning private and social incentives, and when these gains are higher than symmetric firm models.[4]

To understand efficiency in general equilibrium, we examine resource allocation in the standard setting of a monopolistically competitive industry with heterogeneous firm productivity and free entry (e.g. Melitz 2003). We begin our analysis by considering constant elasticity of substitution (CES) demand. In this setting, we show market allocations are efficient, despite differences in firm productivity. This is striking, as it requires the market to induce optimal resource allocations across aggregate variety, quantity and productivity. As in symmetric firm models, there are two sources of potential inefficiency: the inability of firms to appropriate the full consumer surplus and to account for business stealing from other firms. CES demand uniquely ensures these two externalities exactly offset each other. Firm heterogeneity does not introduce any new distortions because the magnitude of these externalities does not vary across firms. Each firm however charges a price higher than its average cost of production. When productivity differs, 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, and positive profits are efficient. Therefore, the market implements the first-best allocation and laissez faire industrial policy is optimal.[5]

What induces market efficiency and how broadly does this result hold? We generalize the demand structure to the variable elasticity of substitution form of Dixit and Stiglitz (1977), which provides a rich setting for a wide range of market outcomes (Vives 2001; Zhelobodko, Kokovin, Parenti and Thisse 2012). When demand elasticity varies with quantity and firms vary in productivity, markups vary within a market. This accounts for the stylized facts that firms are rarely equally productive and markups are unlikely to be constant.[6] Introducing this empirically relevant feature of variable elasticities turns out to be crucial in understanding distortions. When elasticities vary, firms differ in market power and market allocations reflect the distortions of imperfect competition. Nonetheless, we show the market maximizes real revenues. This is similar to perfect competition models, but now market power implies private benefits to firms

[4]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.

[5]Melitz (2003) considers both variable and fixed costs of exporting. The open Melitz economy is efficient, even with 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.

[6]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, Haltiwanger and Syverson (2008) show that "profitability" rather than productivity is more important for firm selection, suggesting a role for richer demand specifications. For further discussion, see Melitz and Trefler (2012).

are perfectly aligned with social benefits only under CES demand. More generally, the appropriability and business stealing effects need not exactly offset each other, and firm heterogeneity introduces a new source of potential inefficiency. When firm differ in productivity, entry of an additional variety has different business stealing effects across the entire distribution of firms and induces distortions relative to optimal allocations.

The pattern of distortions is determined by two elasticities: the demand elasticity, which measures market incentives through markups, and the elasticity of utility, which measures social incentives through a firm's contribution to welfare. We show the way in which these incentives differ characterizes the precise nature of misallocations. This also yields two new insights relating productivity differences to misallocations. First, differences in market power across firms imply misallocations are not uniform: some firms over-produce while others under-produce within the same market. For instance, the market may favor excess entry of low productivity firms, thereby imposing an externality on high productivity firms who end up producing too little. Second, differences in market power impact economy-wide outcomes. The distribution of markups affects ex ante profitability, and therefore the economy-wide trade-off between aggregate quantity and variety. This is in sharp contrast to symmetric firm markets, where markups (or demand elasticities) do not matter for misallocations, as emphasized by Dixit and Stiglitz (1977) and Vives (2001). Differences in productivity underline the importance of demand elasticity for allocational efficiency, and complement the message of Weyl and Fabinger (2012) and Parenti et al. (2014) that richer demand systems enable a better understanding of market outcomes.

As misallocations vary by firm productivity, one potential policy option that does not require firm-level information is international integration. The idea of introducing foreign competition to improve efficiency goes back at least to Melvin and Warne (1973). We show that market integration always provides welfare gains when private and social incentives are aligned, which again is characterized by the demand elasticity and the elasticity of utility. This result ties the Helpman-Krugman characterization of gains from trade to the welfare approach of Spence-Dixit-Stiglitz. Symmetric firm models with CES demand provide a lower bound for the welfare gains from integration. Gains from trade under aligned preferences are higher due to selection of the right distribution of firms and and allocation of the right quantities across firms. International integration therefore has the potential to reduce the distortions of imperfectly competitive markets.

Related Work. Our paper is related to work on firm behavior and welfare in industrial organization and international economics. As mentioned earlier, the trade-off between quantity and variety occupies a prominent place in the study of imperfect competition. We contribute to this literature by studying these issues in markets where productivity differences are important. To highlight the potential scope of market imperfections, we consider variable elasticity of substitution (VES) demand. In contemporaneous work, Zhelobodko et al. (2012) demonstrate the richness and tractability of VES market outcomes under various assumptions such as multiple sectors and vertical differentiation.[7] The focus on richer demand systems is similar to Weyl and Fabinger (2012) who characterize several industrial organization results in terms of pass-through rates. Unlike these papers, we examine the efficiency of market allocations, so our findings depend on both the elasticity of utility and the demand elasticity. To the best of our knowledge, this is the first paper to show market outcomes with heterogeneous firms are first-best under CES demand.[8]

The findings of our paper are also related to a tradition of work on welfare gains from trade. Helpman and Krugman (1985) and Dixit and Norman (1988) examine when trade is beneficial under imperfect competition. We generalize their finding and link it to model primitives of demand elasticities, providing new results even in the symmetric firm literature. In recent influential work, Arkolakis et al. (2012a,b) show richer models of firm heterogeneity and variable markups are needed for these microfoundations to affect the mapping between trade data and welfare gains from trade. In line with this insight, we generalize the demand structure and show that firm heterogeneity and variable markups matter for both welfare gains and allocational efficiency.[9] As in Melitz and Redding (2013), we find that the cost distribution matters for the magnitude of welfare gains from integration. Building on Bernard, Eaton, Jensen and Kortum (2003), de Blas and Russ (2010) also examine the role of variable markups in welfare gains but do not consider efficiency. We follow the direction of Tybout (2003) and Katayama, Lu and Tybout (2009) which suggest the need to map productivity gains to welfare and optimal

[7]While VES utility does not include the quadratic utility of Melitz and Ottaviano (2008) and the translog utility of Feenstra (2003), Zhelobodko et al. show it captures the qualitative features of market outcomes under these forms of non-additive utility.

[8]We consider this to be the proof of a folk theorem which has been "in the air." Matsuyama (1995) and Bilbiie, Ghironi and Melitz (2006) find the market equilibrium with symmetric firms is socially optimal only when preferences are CES. Epifani and Gancia (2011) generalize this to multiple sectors while Eckel (2008) examines efficiency when firms affect the price index. 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 productivity increases are offset by reductions in variety. We provide an analytical treatment to show the market equilibrium implements the unconstrained social optimum. Helpman, Itskhoki and Redding (2011) consider the constrained social optimum. Their approach differs because the homogeneous good fixes the marginal utility of income. Our work is closest to Feenstra and Kee who focus on the CES case. Considering 48 countries exporting to the US in 1980-2000, they also estimate that rise in export variety accounts for an average 3.3 per cent rise in productivity and GDP for the exporting country.

[9]For instance, linear VES demand and Pareto cost draws fit the gravity model, but firm heterogeneity still matters for market efficiency. More generally, VES demand is not nested in the Arkolakis et al. models and does not satisfy a log-linear relation between import shares and welfare gains, as illustrated in the Online Appendix.

policies.

The paper is organized as follows. Section 2 recaps the standard monopolistic competition framework with firm heterogeneity. Section 3 contrasts efficiency of CES demand with inefficiency of VES demand and Section 4 characterizes the distortions in resource allocation. Section 5 examines the welfare gains from integration and Section 6 concludes.

2 Model

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

2.1 Consumers

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

An economy consists of a mass L of identical workers, each endowed with one unit of labor and facing a wage rate w normalized to one. Workers have identical preferences for a differentiated good. If the subset of horizontally differentiated varieties available to them is $[0,N]$, and the worker chooses quantity q_i of variety $i \in [0,N]$ her utility takes the general VES form:

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

where $u(\cdot)$ satisfies Assumption 1 below.

Assumption 1. *Utility Restrictions.*

1. *(Regularity) u is strictly increasing, concave, four times continuously differentiable, satisfies inada conditions and* $u(0) = 0$.

The concavity of u ensures consumers love variety and prefer to spread their consumption over all available varieties. Here $u(q_i)$ denotes utility from an individual variety i. Under CES preferences, $u(q_i) = q_i^\rho$ as specified in Dixit-Stiglitz and Krugman (1980).[10]

Given prices p_i for the varieties, each worker maximizes her utility subject to her budget constraint. VES preferences induce an inverse demand $p(q_i) = u'(q_i)/\delta$ for variety i where δ is the consumer's budget multiplier. As u is strictly increasing and concave, for any fixed price vector the consumer's maximization problem is concave. The necessary condition which determines the inverse demand is sufficient, and has a solution provided inada conditions on u.[11] Multiplying both sides of the inverse demand by q_i and aggregating over all i, the budget multiplier is $\delta = \int_0^N u'(q_i) \cdot q_i di$. The consumer budget multiplier δ will act as a demand shifter and the inverse demand will inherit the properties of the marginal utility $u'(q_i)$. In particular, the inverse demand elasticity $|d\ln p_i/d\ln q_i|$ equals the elasticity of marginal utility $\mu(q_i) \equiv |q_i u''(q_i)/u'(q_i)|$, which enables us to characterize market allocations in terms of demand primitives. Under CES preferences, the elasticity of marginal utility is constant and the inverse demand elasticity does not respond to consumption ($|d\ln p_i/d\ln q_i| = \mu(q_i) = 1-\rho$). When $\mu'(q_i) > 0$, the inverse demand of a variety becomes more elastic as its consumption increases. The opposite holds for $\mu'(q_i) < 0$, where the demand for a variety becomes less elastic as its price rises. Zhelobodko et al. (2012) show that the elasticity of marginal utility $\mu(q_i)$ can also be interpreted in terms of substitution across varieties. For symmetric consumption levels ($q_i = q$), this elasticity equals the inverse of the elasticity of substitution between any two varieties. For $\mu'(q) > 0$, higher consumption per variety or fewer varieties for a given total quantity, induces a lower elasticity of substitution between varieties. Consumers perceive varieties as being less differentiated when they consume more, but this relationship does not carry over to heterogeneous consumption levels.

The inverse demand elasticity summarizes market demand, and will enable a characterization of market outcomes. A policymaker maximizes utility, and is not concerned with market prices. Therefore, we define the elasticity of utility $\varepsilon(q_i) \equiv u'(q_i)q_i/u(q_i)$, which will enable a characterization of optimal allocations. For symmetric consumption levels, Vives (2001) points out that $1-\varepsilon(q)$ is the degree of preference for variety as it measures the utility gain from adding a variety, holding quantity per firm fixed. To arrive at an analogue of the discrete good case, consider a consumer who ceases to purchase a basket of varieties $[0,\alpha]$. The consumer

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

[11] Additional assumptions to guarantee existence and uniqueness of the market equilibrium are in a separate note available online (Dhingra and Morrow 2016b). Utility functions not satisfying inada conditions are permissible but may require parametric restrictions to ensure existence.

loses an average utility of $\int_0^\alpha u(q_i)\,di/\alpha$ per variety not purchased and saves an average income of $\int_0^\alpha p_i q_i di/\alpha$ per variety. The savings can be used to increase consumption of all other varieties proportionally by $\int_0^\alpha p_i q_i di/\int_\alpha^N p_i q_i di$, leading to a rise in average utility per variety not purchased of (using $p_i = u'(q_i)/\delta$)

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

Letting α approach zero gives us an expression for how $1-\varepsilon$ measures the net welfare gain of purchasing additional variety (here, variety 0): a welfare gain of $u(q_0)$ at a welfare cost of $\delta p_0 q_0 = u'(q_0)q_0$ by proportionally consuming less of other varieties. Thus $1-\varepsilon(q_0) = (u(q_0) - u'(q_0)q_0)/u(q_0)$ denotes the welfare contribution of variety relative to quantity.

We summarize the assumptions for a well-defined consumer budgeting problem in Assumption 2 below.

Assumption 2. *Consumer Regularity Conditions.*

1. *(Bounded Elasticities) The elasticity of utility $\varepsilon(q)$ and elasticity of marginal utility $\mu(q)$ are bounded below by $m > 0$ and above by $1-m < 1$.*
2. *(Non-satiation)* $\sup_q \left\{U(N,q) : \int_0^N p_i q_i di = 1\right\} < \infty$.
3. *(Bounded Expenditure)* $\int_0^N p_i (u')^{-1}\left(\delta^{\text{finite}} p_i\right) di < \infty$ for some $\delta^{\text{finite}} > 0$.

Assumption 2.1 maintains boundedness between aggregate costs, revenues and welfare. Assumption 2.2 is automatically satisfied if u is bounded, but more broadly is an assumption that the prices faced by a consumer do not allow consumers to attain infinite welfare conditional on the distribution of prices, for instance if many goods have prices close to zero. Assumption 2.3 is a condition that guarantees the prices presented to consumers imply finite expenditure. Assumptions 2.2 and 2.3 will be ensured in equilibrium once firm behavior is considered. We turn to the firms' production and entry decisions in the next sub-section.

2.2 Firms

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

After entry, should a firm produce, it incurs a fixed cost of production $f > 0$. Profit maximization implies firms produce if they can earn non-negative profits net of the fixed costs of production. A firm with cost draw c chooses its quantity $q(c)$ to $\max_{q(c)}[p(q(c)) - c]q(c)L$ and $q(c) > 0$ if $\pi(c) = \max_{q(c)}[p(q(c)) - c]q(c)L - f > 0$. To ensure the firm's quantity FOC is optimal, we assume marginal revenue is strictly decreasing in quantity. Assumption 3 below summarizes the conditions for a well-behaved profit maximization problem.

Assumption 3. *Firm Regularity Conditions.*

1. *(Decreasing Marginal Revenue) Real revenues $u'(q) \cdot q$ are strictly concave in quantity.*[12]
2. *(Bounded Costs)* $\int_0^\infty c \cdot (u')^{-1}\left(\delta^{\text{finite}} c\right) dG(c) < \infty$ for some $\delta^{\text{finite}} > 0$.

Assumption 3.1 guarantees the monopolist's FOC is optimal, the quantity choice is determined by the equality of marginal revenue and marginal cost, and that quantities are uniquely defined for any positive, finite δ. Assumption 3.2 is a condition that guarantees the distribution of costs in conjunction with demand allows for finite resource usage by a unit mass of firms.

A firm chooses its quantity to equate marginal revenue and marginal cost ($p + q \cdot u''(q)/\delta = c$), and concavity of the firm problem ensures low cost firms supply higher quantities and charge lower prices. The markup charged by a firm with cost draw c is $(p(c) - c)/p(c) = -q(c)u''(q(c))/u'(q(c))$. This shows that the elasticity of marginal utility $\mu(q)$ summarizes the markup:

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

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

2.3 Market Equilibrium

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

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

[12] Inada conditions for revenue are implied by Assumptions 1 and 2.1 since $[u'(q) \cdot q]' = u'(q)[1 - \mu(q)]$.

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

We will compare the free entry market equilibrium with the socially optimal allocation. In a separate note, we show that the assumptions defined in terms of model primitives, Assumptions 2.1 and 3, ensure there is a unique market equilibrium and the quantities produced are continuously differentiable in costs. In the remainder of this Section, we state the social planner's problem and then proceed to a comparison of the market and optimal allocations.

2.4 Social Optimum

A policymaker maximizes individual welfare U as given in Equation (1) by choosing the mass of entrants, quantities and types of firms that produce.[13] The policymaker can choose any allocation of resources that does not exceed the total resources in the economy. She faces the same entry process as for the market: a sunk entry cost f_e must be paid to get a unit cost draw from $G(c)$. Fixed costs of production imply that the policymaker chooses zero quantities for varieties above a cost threshold. Therefore, all optimal allocation decisions can be summarized by quantity $q(c)$, potential variety M_e and a productivity cutoff c_d. Then the policymaker's problem is to choose $q(c)$, c_d and M_e to

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

3 Market Efficiency

Having described an economy consisting of heterogeneous, imperfectly competitive firms, we now examine efficiency of market allocations. This Section starts with a discussion of the potential externalities in the market and efficiency under CES demand. Then we explain market inefficiency under VES demand.

3.1 Market and Optimal Allocations

Outside of cases in which imperfect competition leads to competitive outcomes with zero profits, one would expect the coexistence of positive markups and positive ex post profits to indicate

[13] Free entry implies zero expected profits, so the focus is on consumer welfare.

inefficiency through loss of consumer surplus. Nonetheless, we show that CES demand under firm heterogeneity exhibits positive markups and ex post profits for surviving firms, yet it is allocationally efficient. However, this is a special case. Private incentives are not aligned with optimal production patterns for any VES demand structure except CES.

Proposition 1 shows the market provides the first-best quantity, variety and productivity. The proof of Proposition 1 differs from symmetric firm monopolistic competition results because optimal quantity varies non-trivially with unit cost, variety and cutoff productivity. The main finding is that laissez faire industrial policy is optimal under CES demand.

Proposition 1. *Under CES demand $u(q) = q^{\rho}$ for $0 < \rho < 1$ and Assumptions 1, 2 and 3, there is a unique market equilibrium at which quantities produced are continuously differentiable in costs and it is socially optimal.*

Proof. See Appendix. □

Proposition 1 shows that the market allocation is optimal under CES demand and we now contrast the market allocation across symmetric and heterogeneous firms. When firms are symmetric, resource allocation reflects average cost pricing. Firms charge positive markups which result in lower quantities than those implied by marginal cost pricing. Even though firms do not charge marginal costs, their market price (and hence marginal utility) is proportional to marginal cost because markups are constant. This ensures proportionate reductions in quantity from the level that would be observed under marginal cost pricing (Baumol and Bradford 1970). These reduced quantity levels are efficient because the marginal utility of income adjusts to ensure that the ratio of marginal utility to marginal cost of a variety coincides with the social value of labor ($u'(q)/c = \delta/(1-\mu) = \lambda$). Free entry equates price to the average cost of production, and the markup exactly finances the fixed cost of an additional variety. The market therefore induces an efficient allocation.

With heterogeneous firms, markups continue to be constant and marginal utility is proportional to marginal cost. One might infer enforcing average cost pricing across different firms would induce an efficient allocation, as in symmetric firm models. But average cost pricing is too low to compensate firms because it will not cover ex ante entry costs. The market ensures prices above average costs at a level that internalizes the losses faced by exiting firms. Entry is at optimal levels that fix $p(c_d)$, thereby fixing absolute prices to optimal levels. Post entry, surviving firms charge prices higher than average costs ($p(c) \geq [cq(c) + f/L]/q(c)$). Their ex post profits are positive but the markups exactly compensate them for the possibility of paying f_e to enter and then being too unproductive to survive.

The way in which CES preferences cause firms to optimally internalize aggregate economic conditions can be made clear through a variety-specific explanation. The elasticity of utility $\varepsilon(q) \equiv u'(q) \cdot q/u(q)$ can be used to define a "social markup" $1-\varepsilon(q)$. We term $1-\varepsilon(q)$ the social markup because it denotes the utility from consumption of a variety net of its resource cost. At the optimal allocation, the multiplier λ encapsulates the social value of labor and the social surplus from a variety is $u(q)-\lambda cq$. At the optimal quantity, $u'(q(c)) = \lambda c$ and the social markup is

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

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

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

In contrast, the incentives that firms face in the market are based on the private markup $\mu(q) = (p(q)-c)/p(q)$, and firms solve:

$$\max_q (p(q)q - cq)L - f = \frac{\mu(q^{\text{mkt}}(c))}{1-\mu(q^{\text{mkt}}(c))} cq^{\text{mkt}}(c)L - f.$$

Since ε and μ depend only on the primitive $u(q)$, we can examine what demand structures would make the economy optimally select firms. Clearly, if private markups $\mu(q)$ coincide with social markups $1-\varepsilon(q)$, "profits" will be the same at every unit cost. Examining CES demand, we see precisely that $\mu(q) = 1-\varepsilon(q)$ for all q. Thus, CES demand incentivizes exactly the right firms to produce. Since the optimal set of firms produce under CES demand, and private and social profits are the same, market entry will also be optimal. As entry M_e and the cost cutoff c_d are optimal, the competition between firms aligns the budget multiplier δ to ensure optimal quantities.

Efficiency of the market equilibrium in our framework is tied to CES demand. While comparing private and social markups provides a simple way to understand efficiency, this variety-specific comparison does not explain the general equilibrium forces that induce efficiency. Under symmetric firms, Mankiw and Whinston (1986) show there are two externalities that arise in the market. First, firms cannot capture the entire surplus generated by their production, and this lack of appropriability discourages firm entry. This is summarized by the elasticity of utility which measures the proportion of utility from a variety not captured by its real revenue ($1-\varepsilon(q) = 1-u'(q)q/u(q)$). Second, firms do not internalize the downward pressure imposed by their production on prices of other firms, and this business stealing effect tends to encourage too much entry. This externality is summarized by the inverse demand elasticity $\mu(q)$. Under CES demand, the market allocation is efficient because the appropriability externality balances the business stealing externality $(1-\varepsilon-\mu=0)$ for each variety, and there is no incentive to deviate from optimal entry (Grossman and Helpman 1993). When firms differ in productivity, CES demand continues to ensure the market allocates resources optimally because each variety has the same levels of appropriability and business stealing. These externalities exactly counteract each other and there are no new distortions due to firm heterogeneity. More generally, the appropriability externality and the business stealing externality differ from each other and the magnitude of these effects vary across firms. To highlight this, we consider the general class of VES demand specified in Equation (1). Direct 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.

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

Proof. Online Appendix. □

Under general VES demand, market allocations are not efficient and do not maximize individual welfare. Lemma 1 shows that the market instead maximizes aggregate real revenue generated in the economy. Defining the real revenue per variety as $u'(q)q$, aggregate real revenue $(M_e \int_0^{c_d} u'(q(c)) \cdot q(c) dG)$ is maximized for all VES demand functions with positive and decreasing marginal revenues, as stated below.

Lemma 1. *Under VES demand satisfying Assumptions 1, 2 and 3, the market maximizes aggregate real revenue $M_e \int_0^{c_d} u'(q(c)) \cdot q(c) dG$ subject to the resource constraint of the economy: $L \geq M_e \left\{ \int_0^{c_d} [cq(c)L + f] dG + f_e \right\}$ where c_d is the cost cutoff from the market equilibrium.*

Proof. See Appendix. □

Lemma 1 shows decentralized profit maximization coincides with centralized revenue maximization. The intuition is analogous to perfect competition where the economy can be subdivided into smaller replications which must reward factors identically and therefore maximize revenues to pay factors identically. Under VES, the economy can be examined as subdivisions of ex ante entrants. Free entry and additive utility across varieties imply than if any fraction of labor was used by a subset of entrants which did not maximize revenue as a group, then the total wage bill they could pay (equal to revenues) would be lower than for the same fraction of labor elsewhere in the economy, violating labor market clearing.

This result shows that the market and optimal allocations are generally not aligned under VES demand. The market and optimal allocations are solutions to:

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

For CES demand, $u(q) = q^\rho$ while $u'(q)q = \rho q^\rho$ implying revenue maximization is perfectly aligned with welfare maximization. The CES result is therefore a limiting case of allocations under VES demand. Outside of CES, quantities produced by firms are too low or too high and in general equilibrium, this implies productivity of operating firms is also too low or too high. Market quantity, variety and productivity reflect distortions of imperfect competition. This leads us to an examination of the distribution of misallocations under VES demand.

4 Market Distortions and Variable Elasticities

As discussed earlier, the market equilibrium has two externalities: the appropriability externality measured by $(1 - \varepsilon(q))$ and the business stealing externality measured by $\mu(q)$. When private and social markups vary with quantity, these two externalities do not offset each other exactly and the market misallocates resources. Dixit and Stiglitz (1977) examine when the market induces optimal entry of symmetric varieties, and find that the bias in market allocation is determined by how the elasticity of utility varies with quantity $(1 - \varepsilon(q))'$. When firms differ in productivity, the business stealing effect varies across varieties, and we show that the inverse demand elasticity $\mu'(q)$ matters for the bias in market allocations. To explain the misallocations, this Section starts with a discussion of the relationship between markups and quantity ($\mu'(q)$ and $(1 - \varepsilon(q))'$). Then we re-state the link between markup variation and misallocations in a symmetric firm setting. Finally, we characterize how the market allocates resources relative to the social optimum under firm heterogeneity, and discuss the robustness of these results to different modeling assumptions.

4.1 Markup and Quantity Patterns

We will show that the relationship between markups and quantity characterizes distortions. When $\mu'(q) > 0$, private 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), so low cost firms have both high q and high markups. When $\mu'(q) < 0$, small "boutique" firms charge higher markups. Similarly, the sign of $(1-\varepsilon(q))'$ determines how social markups vary with quantity. For CES demand, private and social markups are constant ($\mu' = 0$, $(1-\varepsilon)' = 0$).

The empirical relationship between markups and quantities is largely in line with increasing private markups $\mu'(q) > 0$, though decreasing markups are also a theoretical possibility. De Loecker et al. (2012); De Loecker and Warzynski (2012); Dhyne et al. (2011) find markups are positively correlated with firm productivity of manufacturing firms, implying $\mu'(q) > 0$. The bulk of studies that infer markups from the price response to exchange rate fluctuations also find evidence for $\mu'(q) > 0$ (Goldberg and Knetter 1997). In early work, Klette (1999) however shows Norwegian firms with higher markups tend to have lower productivity.[14]

The empirical literature largely finds increasing firm markups, but social markups are rarely observable. We therefore discuss the theoretical implications of different signs for $(1-\varepsilon)'$. For this purpose, it is useful to define preferences according to how private and social markups vary with quantity. Definition 1 below characterizes preferences as aligned when private and social markups move in the same direction.

Definition 1. Private and social incentives are *aligned* when μ' and $(1-\varepsilon)'$ have the same sign.

Commonly-used utility functions exhibit aligned preferences. For instance, $(1-\varepsilon)' > 0$ whenever $\mu' > 0$ in the HARA class. 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 and[15] HARA.[16]

Table 1: Private and Social Markups for Common Utility Forms

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

Conversely, incentives are *misaligned* when μ' and $(1-\varepsilon)'$ have different signs. There are reasons to believe that misaligned preferences are less appealing for theoretical work. The most commonly used misaligned preferences are the generalized Dixit-Stiglitz CES preferences

[14] In a series of influential papers, Roberts and Supina (1996, 2001) show that six of the thirteen manufactured products in their US data exhibit increasing price-cost margins, four products have decreasing price-cost margins and two products show no significant variation. But these studies focus on products that are relatively homogeneous across manufacturers, such as white pan bread and ready-mixed concrete.

[15] The expo-power utility was proposed by Saha (1993) and recently used by Holt and Laury (2002) and Post, Van den Assem, Baltussen and Thaler (2008) to model risk aversion empirically.

[16] The parameter restrictions are $\rho \in (0,1)$ and $\alpha > q/(\rho-1)$ for HARA.

$u(q) = (q+\alpha)^\rho$ for $\alpha > -q$ but these preferences are not continuous at zero when they are appropriately normalized to ensure $u(0) = 0$. It is also unclear whether well-behaved preferences can be misaligned across all quantity levels. Vives (2001) shows aligned preferences also have the advantage that the elasticity of $1-\varepsilon$ equals the elasticity of μ in the limit as q approaches zero under a relatively mild assumption. For these reasons, we focus on the case of aligned preferences and especially on increasing private and social markups which are empirically relevant.

4.2 Misallocations under Symmetric Firms

Dixit and Stiglitz examine how the market allocation deviates from the optimal allocation. They find that the elasticity of utility determines the bias in production and entry. We state their result below and discuss how productivity differences affect distortions subsequently.

Proposition 3. *Under symmetric firms, the pattern of misallocation is as follows:*
1. If $(1-\varepsilon)' < 0$*, market quantities are too high and market entry is too low.*
2. If $(1-\varepsilon)' > 0$*, market quantities are too low and market entry is too high.*

Proof. Dixit and Stiglitz (1977). □

Variation in the elasticity of utility summarizes the difference between the lack of appropriability and business stealing because $\varepsilon' q/\varepsilon = 1-\varepsilon-\mu$. When $(1-\varepsilon)' > 0$, the business stealing externality outweighs the appropriability externality. Firms ignore the negative effect of entry on prices and the market provides too much variety. When $(1-\varepsilon)' < 0$, the business stealing externality is smaller and the market provides too little variety. Under symmetric firms, the two externalities are the same across all firms and the variation in firm markups $\mu'(q)$ does not affect the bias in market allocations.

The symmetric firm case simplifies the analysis of misallocations as the tradeoff is between two decisions: quantity and entry. In contrast, determining misallocations across heterogeneous firms is less obvious because quantities vary by firm productivity, and this variation depends on entry and selection. Further, the business stealing effect varies across firms and depends on the distribution of markups. The next sub-section explains these misallocations for heterogeneous firms. Examining misallocations across the entire distribution of firms reveals two substantive results. First, as we might expect, the misallocation of resources across firms differs by productivity. An interesting finding is that this heterogeneity in misallocation can be severe enough that some firms over-produce while others under-produce. For example, as we will show below, when $\mu' > 0$ and $(1-\varepsilon)' > 0$, excess production by small firms imposes an externality on large firms. Large firms produce below their optimal scale and too many small firms enter the market. In this case, the market diverts resources away from large firms towards small firms. Second, accounting for firm heterogeneity shows that both the elasticity of utility and the inverse demand elasticity determine resource misallocations. When firms are symmetric, only the elasticity of utility determines misallocations and the inverse demand elasticity does not matter (Proposition 3). The presence of productivity differences across firms fundamentally changes the qualitative analysis. When markups vary, firms with different productivity levels charge different markups and steal business at different rates across firms. Therefore, firm heterogeneity and variable markups alter the standard policy rules for correcting misallocation of resources.

4.3 Quantity, Selection and Entry Distortions

We now characterize the misallocations by demand characteristics. The distortions in quantity, productivity and entry are discussed in turn. The distortions depend on both μ' and $(1-\varepsilon)'$, firm heterogeneity reveals new distortions in market outcomes.

4.3.1 Quantity Bias

Quantity distortions differ across firms, and whether firms over-produce or under-produce depends on their productivity. The relationship between market and optimal quantities is fixed by FOCs for revenue maximization and welfare maximization. The market chooses $[1-\mu(q^{\text{mkt}})]u'(q^{\text{mkt}}) = \delta c$, while the optimal quantity is given by $u'(q^{\text{opt}}) = \lambda c$. Therefore, the relationship of market and optimal quantities is

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

The ratio of real revenue to welfare δ/λ depends on entry, productivity and the distribution of quantities. It summarizes the industry-wide distortions through the lack of appropriability and business stealing across all varieties. The variety-specific externality arises from different rates of business stealing which is captured by $\mu(q^{\text{mkt}}(c))$.

When incentives are *aligned*, the gap between the market and social cost of resources (δ and λ) is small enough that quantities are not uniformly distorted across all firms. The variety-specific business stealing effect can dominate the average appropriability and business stealing effects, leading to differences in the quantity bias across firms. Quantities are equal for some c^* where $1-\mu\left(q^{\text{mkt}}(c^*)\right) = \delta/\lambda$. For all other varieties, quantities are still distorted. When $\mu', (1-\varepsilon)' > 0$, market production is biased towards high cost firms ($q^{\text{mkt}} < q^{\text{opt}}$ for low c and

$q^{\mathrm{mkt}} > q^{\mathrm{opt}}$ for high c). The market shifts business away from low cost firms and over-rewards high cost firms. When $\mu', (1-\varepsilon)' < 0$, the bias is reversed and low cost firms over-produce. Therefore, when private and social markups are aligned, whether the market under or over produces depends on a firm's costs. Proposition 4 summarizes the bias in market quantities.

Proposition 4. *When preferences are aligned and* $\inf_q \varepsilon(q) > 0$, $q^{\mathrm{mkt}}(c)$ *and* $q^{\mathrm{opt}}(c)$ *have a unique crossing* c^* *(perhaps beyond market and optimal cost cutoffs).*
3. If $(1-\varepsilon)' > 0$ *and* $\mu' > 0$ *,* $q^{\mathrm{mkt}}(c) < q^{\mathrm{opt}}(c)$ *for* $c < c^*$ *and* $q^{\mathrm{mkt}}(c) > q^{\mathrm{opt}}(c)$ *for* $c > c^*$.
4. If $(1-\varepsilon)' < 0$ *and* $\mu' < 0$, $q^{\mathrm{mkt}}(c) > q^{\mathrm{opt}}(c)$ *for* $c < c^*$ *and* $q^{\mathrm{mkt}}(c) < q^{\mathrm{opt}}(c)$ *for* $c > c^*$.

4.3.2 Selection Bias

The distortion in firm selection is determined by the relation between the elasticity of utility and quantity. Proposition 5 shows that market productivity is either too low or high, depending on whether social markups are increasing or decreasing. We use this result now to depict the pattern of misallocation graphically, and discuss the result further below.

Proposition 5. *Market selection is too low or high, as follows:*
1. If $(1-\varepsilon)' > 0$*, market selection is too low:* $c_d^{\mathrm{mkt}} > c_d^{\mathrm{opt}}$.
2. If $(1-\varepsilon)' < 0$*, market selection is too high:* $c_d^{\mathrm{mkt}} < c_d^{\mathrm{opt}}$.

While Proposition 5 follows from a general equilibrium analysis, the decision to introduce a marginal variety can be intuitively explained as follows. Under increasing social markups $(1-\varepsilon)' > 0$, the lack of appropriability of a marginal variety is lower than its business stealing effect. This encourages production of the marginal variety and the cost cutoff in the market is too high. Although the marginal variety steals business at different rates across varieties, its impact on reallocation of business across varieties is small and the bias in the cost cutoff is determined by the elasticity of utility.

Propositions 4 and 5 show the market misallocates resources across firms, and variable demand elasticities characterize the pattern of these misallocations. Figure 1 illustrates the bias in firm-level production for aligned preferences when markups increase in quantity. For ease of reference, Table 2 summarizes the misallocations by demand characteristics.[17]

[17] Table 2 characterizes the qualitative role of demand elasticities in misallocations. Using a quantitative measure of distortions reiterates their importance. The loss from misallocations can be summarized by the difference between social and market "profits", evaluated at optimal allocations. This measure consists of the difference between average social markup and average private markup $(1-\bar{\varepsilon}-\bar{\mu})$, and the covariance between social and private markups $\mathrm{Cov}(1-\varepsilon, \mu)$. The covariance component shows that the distribution of markups matters for quantifying distortions, except when firms are symmetric or markups are constant (leading to zero covariance).

Figure 1: Bias in Firm Production for Aligned Preferences with Increasing Markups
$\mu', (1-\varepsilon)' > 0$

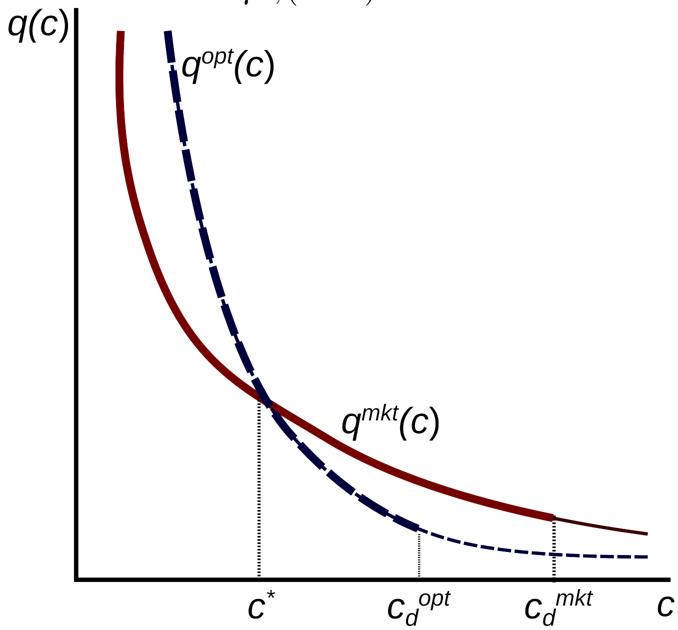

Table 2: Distortions for Aligned Preferences by Demand Elasticities

Increasing Markups $\mu', (1-\varepsilon)' > 0$	Decreasing Markups $\mu', (1-\varepsilon)' < 0$
Quantities High-Cost Skewed: $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c < c^*$ $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c > c^*$	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: $c_d^{\text{mkt}} > c_d^{\text{opt}}$	Productivity Too High: $c_d^{\text{mkt}} < c_d^{\text{opt}}$

A comparison of the mass of entrants in the market and the optimum is generally hard to make. Unlike the marginal variety, entry has sizable business stealing effects. The two externalities of appropriability and business stealing move in opposite directions, and the bias in potential entry M_e and available variety $M_eG(c_d)$ cannot be determined without further information on demand and cost parameters. The net effect of the externalities depends on the relative magnitudes of demand and cost parameters including the cost distribution $G(c)$. For instance, Nocco, Ottaviano and Salto (2013) find that the mass of firms cannot be unambiguously ranked even when demand is linear and the cost distribution is Pareto. While firm heterogeneity makes entry distortions dependent on the cost distribution, the bias in quantity and selection can be unambiguously inferred from the demand-side elasticities.

4.4 Extensions of the Basic Framework

As many different fields of economics (such as macroeconomics and urban economics) use monopolistically competitive models, we extend our basic framework to different modeling assumptions used in these fields to discuss the robustness of CES efficiency and misallocations under VES demand. The main finding is that allowing the costs of production of a variety to vary with its scale of production does not change the market distortion results of Sections 3 and 4. To account for non-constant marginal costs, let the variable cost of production be $\omega(q) \cdot cq$ and assume $2\omega' + \omega'' q > 0$ for all feasible quantities to ensure strict concavity of the firm problem. The market maximizes aggregate revenue under non-constant marginal costs. As firms account for the interdependence between their unit costs and quantity, CES demand ensures the same tradeoff between different externalities and leads to efficient allocations (as shown in an online Appendix). Under VES demand, the bias in quantity and selection are the same as Propositions 4 and 5. Relegating details to an online Appendix, we note that the market distortion results are robust to three other extensions. Allowing for multiple sectors as in Zhelobodko et al. (2012), the market distortion results are valid for a given level of resource allocation to the differentiated goods sector. Further, the results are robust to a firm-specific advertising technology as in Arkolakis (2010) and to the CES-Benassy class of preferences (Benassy (1996), Alessandria and Choi (2007)).

5 Efficiency and Market Size

Having discussed misallocations, this Section examines gains in welfare and efficiency from integration with world markets. The existence of gains from international trade is one of the "most fundamental results" in economics (Costinot and Rodriguez-Clare (2013)). Increases in market size encourage competition, so we might expect that integration would reduce market power and improve welfare. However, the following insight of Helpman and Krugman (1985) (pp. 179) is relevant:

> Unfortunately imperfect competition, even if takes as sanitized a form as monopolistic competition, does not lead the economy to an optimum. As a result there is no guarantee that expanding the economy's opportunities, through trade or anything else, necessarily leads to a gain. We cannot prove in general that countries gain from trade in the differentiated products model.

Building on this insight, we address two related questions. First, we examine when market expansion provides welfare gains. Having characterized distortions, we first show that welfare gains are related to the demand-side elasticities mentioned earlier. Next, we discuss the role of firm heterogeneity and variable elasticities for quantitative work measuring the welfare gains from international trade.

5.1 Integration, Market Size and Efficiency

We begin with the equivalence between market expansion and trade. Echoing Krugman (1979), an economy can increase its market size by opening to trade with foreign markets. A VES economy of size L_1 that trades freely with countries of sizes $L_2, ..., L_n$ has the same market equilibrium as a single autarkic VES economy of size $L = L_1 + ... + L_n$. Consequently, the market distortions detailed in Section 5 persist in integrated markets. Resource allocation in an integrated market is suboptimal, except under CES demand. When markups vary, marginal revenues do not correspond to marginal utilities so market allocations are not aligned with efficient allocations. This is particularly important when considering trade as a policy option, as it implies that opening to trade may take the economy further from the social optimum. For example, market expansion from trade may induce exit of low productivity firms from the market when it is optimal to keep more low productivity firms with the purpose of preserving variety.

Helpman and Krugman (1985) provide sufficient conditions for welfare gains from trade. They show when productivity and variety do not decline after integration, then there are gains from trade.[18] In terms of primitives, we find integration is always beneficial when preferences are aligned. This is true for any cost distribution, but requires a regularity condition for decreasing private markups ($(\mu + \mu' q/(1-\mu))' \leq 0$). The regularity condition ensures that the marginal revenue is convex whenever firm markups are decreasing in quantity. As market size expands, this implies that the rise in markups from lower per capita quantity is not high enough for firms to scale back on the total quantity that they sell to all consumers in the bigger market. We summarize this in Proposition 6.

Proposition 6. *Market expansion increases welfare when preferences are aligned. (Provided $(\mu + \mu' q/(1-\mu))' \leq 0$ whenever $\mu' < 0$).*

The economic reasoning for Proposition 6 follows from similar responses of the two demand-side elasticities to changes in quantity. An increase in market size increases competition and reduces per capita demand for each variety. When preferences are aligned, demand shifts alter

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

the private and social markups in the same direction. The market therefore incentivizes firms towards the right allocation and provides higher welfare. Building on this result, Bykadorov et al. (2014) show that aligned preferences are necessary and sufficient for welfare gains from trade under symmetric firms and variable marginal costs.

The role of aligned markups in firm survival highlights how trade increases welfare. When aligned markups increase with quantity, a rise in market size forces out the least productive firms. Since social markups are positively correlated with quantity, the least productive firms also contribute relatively little to welfare and their exit is beneficial. When markups decrease with quantity, small "boutique" firms contribute at a higher rate to welfare and are also able to survive after integration by charging higher markups. Integration enables the market to adapt their production in line with social incentives, leading to welfare gains from trade.

5.2 Quantitative Literature on Welfare Gains from Trade

A growing body of work seeks to quantify the welfare gains from trade. New quantitative trade models typically estimate the gains from trade under CES demand. In an influential paper, Arkolakis et al. (2012a) show that welfare in a model with heterogeneous firms can be summarized by two sufficient statistics: the share of expenditure on domestically produced goods and the elasticity of trade with respect to trade costs. As these sufficient statistics are common to heterogeneous and representative firm models, welfare gains estimated from import shares and constant trade elasticities using trade data are the same across heterogeneous and representative firm models. However, the two models only deliver the same estimates for welfare gains when the underlying structural parameters for preferences and technology differ across the models. We use this insight of Melitz and Redding (2013) to explain the relevance of our optimality results for the quantitative literature on the gains from trade.

Melitz and Redding find that the heterogeneous firm model of Melitz provides quantitatively higher gains from trade than an equivalent representative firm model when the structural parameters are the same across these models. As they mention, this can be understood by appealing to the social optimality results for CES demand (Proposition 1). Consider initial equilibria in the heterogeneous and homogeneous firm models that feature identical aggregate statistics and welfare. In the homogeneous firm model, unit cost is exogenously fixed, and hence remains unchanged when the economy opens to trade. In the heterogeneous firm model, the cost distribution changes when the economy opens to trade. In a companion note (Dhingra and Morrow 2016a), we show that the open economy equilibrium with trade frictions is efficient under CES demand. Since the policymaker chooses to change the cost cutoff in an open economy, the open economy market allocation must yield higher welfare than any other feasible allocation (where the unit cost is unchanged). The allocation where the unit cost does not change is identical to the open economy equilibrium in the homogeneous firm model. Therefore the open economy equilibrium in the heterogeneous firm model must yield higher welfare than the open economy equilibrium in the homogeneous firm model. This shows that a quantitative trade model with the same structural parameters across models will provide higher welfare gains in a setting with firm heterogeneity. The optimality of market allocations ensures that firm heterogeneity increases the magnitude of the welfare gains from trade.

Departing from CES preferences, market allocations are no longer optimal. This raises the question of the role played by firm heterogeneity in altering the magnitude of welfare gains from trade. While we do not model trade costs, Proposition 6 shows market expansion through trade provides higher welfare gains when firms differ in productivity. Under aligned preferences and the regularity condition, we show models with firm heterogeneity and increasing markups provide higher welfare gains from trade than representative firm models.

For a given change in real income, the welfare gains from trade depend on the different assumptions on demand and firm costs. Welfare is $U = M_e \int u(q)dG = \delta/\bar{\varepsilon}$ where the average elasticity of utility is $\bar{\varepsilon} \equiv \int \varepsilon u dG / \int u dG$. An increase in market size increases real income at the rate of the average markup ($d\ln\delta/d\ln L = \int \mu pq dG/\int pq dG \equiv \tilde{\mu}$). The change in average elasticity can be decomposed into the change in $\varepsilon(q)$ given $u/\int u dG$, and the change in the weights $u/\int u dG$. Let $x_d \equiv x(q(c_d))$, then the change in the average elasticity of utility is

$$\frac{d\ln\bar{\varepsilon}}{d\ln L} = \int \frac{\varepsilon' u}{\bar{\varepsilon}\int u dG}\frac{d\ln q}{d\ln L}dG + \underbrace{\int \frac{u'\varepsilon - u'\bar{\varepsilon}}{\bar{\varepsilon}\int u dG}\frac{d\ln q}{d\ln L}dG + \frac{u_d}{\bar{\varepsilon}\int u dG}(\varepsilon_d - \bar{\varepsilon})\, c_d g(c_d)\frac{d\ln c_d}{d\ln L}}_{\text{Reallocation across heterogeneous firms}}.$$

The first term is the change in the elasticity of utility due to a fall in quantity per firm, holding fixed the share of each variety in the average elasticity. The second and third terms denote the change in the average elasticity of utility due to a reallocation of resources across heterogeneous firms. Reallocation of resources across firms changes the share of each variety in the average elasticity of utility through $\left(u(q)/\int_0^{c_d} u(q)dG\right)'$. Using this decomposition, we can explain the role of variable elasticities and firm heterogeneity in welfare gains from trade.

For a given change in real income ($d\ln\delta/d\ln L = \tilde{\mu}$), we decompose the gains from trade into gains for a representative firm and gains due to differences in firm productivity. Defining the market outcome of a representative firm as the revenue-weighted average of heterogeneous firms, the gains from trade for a given change in real income are:

$$\frac{d\ln U}{d\ln L} = \underbrace{\tilde{\mu}\int \frac{1-\varepsilon}{\mu}\frac{\varepsilon u}{\bar{\varepsilon}\int u dG}dG}_{\text{CES}} + \underbrace{\tilde{\mu}\int\left(\frac{1-\varepsilon+\mu' q/(1-\mu)}{\mu+\mu' q/(1-\mu)} - \frac{1-\varepsilon}{\mu}\right)\frac{\varepsilon u}{\int \varepsilon u dG}dG}_{\text{VES \& Representative Firm}}$$
$$+ \underbrace{\tilde{\mu}\int \frac{\varepsilon-\bar{\varepsilon}}{\mu+\mu' q/(1-\mu)}\frac{\varepsilon u}{\bar{\varepsilon}\int u dG}dG}_{\text{Quantity Reallocation}} + \underbrace{\frac{u_d}{\int u dG}\frac{c_d g(c_d)}{\bar{\varepsilon}(1-\mu_d)}(\varepsilon_d-\bar{\varepsilon})(\tilde{\mu}-\mu_d)}_{\text{Firm Selection}}$$

The first line contains the gains from trade for a representative firm. The first component is the welfare gain when firm markups are constant and the second component shows how welfare gains change when markups vary with quantity. Under CES demand, the welfare gain is the revenue-weighted average of $1-\varepsilon$. VES demand adds the second component which is positive when markups are increasing and negative when markups are decreasing with quantity.

The second line consists of the gains from trade arising due to differences in firm productivity. The first component of the second line is the welfare gain from changes in *relative* quantities across firms. When firms differ in productivity, market size affects their output levels differently and resources are reallocated across firms. For aligned preferences, quantity reallocation increases the welfare gains from trade under the regularity condition. The second component shows the welfare gains from firm selection. Aligned preferences ensure the market selects the right firms as it expands and leads to higher welfare gains. We summarize this in Proposition 7.

Proposition 7. *Under aligned preferences, the welfare gains from market expansion are higher in heterogeneous firm models, compared to representative firm models. (Provided* $(\mu+\mu' q/(1-\mu))' \leq 0$ *whenever* $\mu' < 0$*).*

When preferences are aligned, the reallocation of resources across heterogeneous firms provides another channel for welfare gains from integration. As most empirical studies are consistent with increasing markups ($\mu' > 0$), structural estimates based on CES demand therefore provide a lower bound $(1-\tilde{\varepsilon})$ for the potential gains from trade. For a given change in real income, accounting for firm heterogeneity and increasing markups would reveal higher welfare gains from trade. The magnitude of these additional gains depends on the markup variation (through $\varepsilon(q(c)) - \varepsilon(q(c_d))$ and $\mu'(q(c))$) and on the productivity distribution (through $g(c_d)$).

6 Conclusion

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

Generalizing to variable elasticities of substitution, firms differ in market power which affects the trade-off between quantity, variety and productivity. Unlike symmetric firm models, the market distortions depend on the elasticity of demand and the elasticity of utility. Under CES demand, these two elasticities are constant and miss out on meaningful trade-offs. When these elasticities vary, the pattern of misallocations depends on how demand elasticities change with quantities, so policy analysis should ascertain these elasticities and take this information into account. While the modeling framework we consider provides a theoretical starting point to understand distortions across firms, enriching the model with market-specific features can yield better policy insights. Neary and Mrazova (2013) and Parenti et al. (2014) provide further generalizations of demand and costs, and Bilbiie et al. (2006) and Opp et al. (2013) consider dynamic misallocations. Future work can also provide guidance on the design of implementable policies to realize further welfare gains.

We focus on international integration as a key policy tool to realize potential gains. Market expansion does not guarantee welfare gains under imperfect competition. As Dixit and Norman (1988) put it, this may seem like a "sad note" on which to end. But we find that integration provides welfare gains when the two demand-side elasticities ensure private and social incentives are aligned. The welfare gains from integration are higher than those obtained in representative firm models with constant elasticities, because market expansion also leads to a better reallocation of resources across heterogeneous firms. Further work might quantify the scope of integration as a tool to improve the performance of imperfectly competitive markets.

References

Alessandria, G. and H. Choi, "Do Sunk Costs of Exporting Matter for Net Export Dynamics?," *The Quarterly Journal of Economics*, 2007, *122* (1), 289–336.

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

_ , _ , D. Donaldson, and A. Rodriguez-Clare, "The Elusive Pro-Competitive Effects of Trade," *Working Paper*, 2012.

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199.

Asplund, M. and V. Nocke, "Firm turnover in imperfectly competitive markets," *The Review of Economic Studies*, 2006, *73* (2).

Atkeson, A. and Burstein, "Innovation, Firm Dynamics, and international Trade," *Journal of Political Economy*, 2010, *118* (3), 433–484.

Baldwin, R. E. and F. Robert-Nicoud, "Trade and growth with heterogeneous firms," *Journal of International Economics*, 2008, *74* (1), 21–34.

Bartelsman, E. J. and M. Doms, "Understanding productivity: Lessons from longitudinal microdata," *Journal of Economic literature*, 2000, *38* (3).

Baumol, W. J. and D. F. Bradford, "Optimal Departures From Marginal Cost Pricing," *The American Economic Review*, 1970, *60* (3), 265–283.

Benassy, J. P., "Taste for variety and optimum production patterns in monopolistic competition," *Economics Letters*, 1996, *52* (1), 41–47.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott, "Firms in International Trade," *The Journal of Economic Perspectives*, 2007, *21* (3), 105–130.

_ , J. Eaton, J. B. Jensen, and S. Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003.

Bilbiie, F. O., F. Ghironi, and M. J. Melitz, "Monopoly power and endogenous variety in dynamic stochastic general equilibrium: distortions and remedies," *manuscript, University of Oxford, Boston College, and Princeton University*, 2006.

Bykadorov, Igor, Alexey Gorn, Sergey Kokovin, and Evgeny Zhelobodko, "Losses from trade in Krugman's model: almost impossible," *Working Paper*, 2014.

Campbell, J. R. and H. A. Hopenhayn, "Market Size Matters," *Journal of Industrial Economics*, 2005, *53* (1), 1–25.

Costinot, Arnaud and Andrés Rodriguez-Clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in "In: Helpman, E.(Ed.), Handbook of international economics" Citeseer 2013.

de Blas, B. and K. Russ, "Understanding Markups in the Open Economy under Bertrand Competition," *NBER Working Papers*, 2010.

Dhingra, Swati and John Morrow, "Efficiency in Heterogeneous Firm Trade Models," *Working Paper*, January 2016.

_ and _ , "Existence and Uniqueness of the Market Equilibrium in a VES Economy," 2016.

Dhyne, Emmanuel, Amil Petrin, and Frederic Warzynski, "Prices, Markups and Quality at the Firm-Product Level," Technical Report, Mimeo, University of Minnesota 2011.

Dixit, A. K. and J. E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

— **and V. Norman**, *Theory of international trade*, Cambridge Univ. Press, 1988.

Eckel, Carsten, "Globalization and specialization," *Journal of International Economics*, May 2008, *75* (1), 219–228.

Epifani, P. and G. Gancia, "Trade, markup heterogeneity and misallocations," *Journal of International Economics*, 2011, *83* (1), 1–13.

Feenstra, R. and H. L. Kee, "Export variety and country productivity: Estimating the monopolistic competition model with endogenous productivity," *Journal of International Economics*, 2008, *74* (2), 500–518.

Feenstra, R. C., "A homothetic utility function for monopolistic competition models, without constant price elasticity," *Economics Letters*, 2003, *78* (1), 79–86.

— , "New Evidence on the Gains from Trade," *Review of World Economics*, 2006, *142* (4), 617–641.

Foster, L., J. C. Haltiwanger, and C. J. Krizan, "Aggregate productivity growth. Lessons from microeconomic evidence," in "New developments in productivity analysis," University of Chicago Press, 2001.

— **, J. Haltiwanger, and C. Syverson**, "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

Goldberg, P. and Michael Knetter, "Goods prices and exchange rates: what have we learned," *Journal of Economic Literature*, 1997, *5862*.

Grossman, Gene M. and Elhanan Helpman, *Innovation and Growth in the Global Economy*, MIT Press, 1993.

Helpman, E. and P. R. Krugman, *Market Structure and Foreign Trade: increasing returns, imperfect competition, and the international economy*, MIT Press, 1985.

— **, O. Itskhoki, and S. J. Redding**, "Trade and Labor Market Outcomes," *NBER Working Paper 16662*, 2011.

Holt, C. A. and S. K. Laury, "Risk aversion and incentive effects," *American Economic Review*, 2002, *92* (5), 1644–1655.

Katayama, H., S. Lu, and J. R. Tybout, "Firm-level productivity studies: illusions and a solution," *International Journal of Industrial Organization*, 2009, *27* (3), 403–413.

Klette, Tor J., "Market power, scale economies and productivity: estimates from a panel of establishment data," *The Journal of Industrial Economics*, 1999, *47* (4), 451–476.

Krugman, P., "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 1979, *9* (4), 469–479.

Krugman, P. R., "Is free trade passé?," *The Journal of Economic Perspectives*, 1987, *1* (2).

Krugman, Paul R., "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70* (5), 950–959.

Loecker, Jan De and Frederic Warzynski, "Markups and Firm-Level Export Status," *The American Economic Review*, 2012, *102* (6), 2437–2471.

_ , Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik, "Prices, markups and trade reform," Technical Report, National Bureau of Economic Research 2012.

Mankiw, N. G. and M. D. Whinston, "Free entry and social inefficiency," *The RAND Journal of Economics*, 1986, pp. 48–58.

Matsuyama, Kiminori, "Complementarities and Cumulative Processes in Models of Monopolistic Competition," *Journal of Economic Literature*, June 1995, *33* (2), 701–729.

Melitz, M. J. and S. J. Redding, "Heterogeneous Firms and Trade," *Handbook of International Trade (commissioned)*, August 2012.

Melitz, Marc and Daniel Trefler, "Gains from Trade when Firms Matter," *Journal of Economic Perspectives*, 2012, *26.*

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

_ and Gianmarco I. P. Ottaviano, "Market Size, Trade, and Productivity," *Review of Economic Studies*, October 2008, *75* (1), 295–316.

_ and Stephen J. Redding, "Firm Heterogeneity and Aggregate Welfare," Technical Report, National Bureau of Economic Research 2013.

Melvin and R. D. Warne, "Monopoly and the theory of international trade," *Journal of International Economics*, 1973, *3* (2), 117–134.

Neary, J. Peter and Monika Mrazova, "Not so demanding: Preference structure, firm behavior, and welfare," Technical Report, University of Oxford, Department of Economics 2013.

Nocco, Antonella, Gianmarco I. P. Ottaviano, and Matteo Salto, "Monopolistic Competition and Optimum Product Selection: Why and How Heterogeneity Matters," *CEP Discussion Paper*, April 2013, *1206.*

Opp, Marcus M., Christine A. Parlour, and Johan Walden, "A Theory of Dynamic Resource Misallocation and Amplification," *Available at SSRN 2049447*, 2013.

Parenti, Mathieu, Philip Ushchev, and Jacques-François Thisse, "Toward a theory of monopolistic competition," 2014.

Pavcnik, N., "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *The Review of Economic Studies*, 2002, *69* (1), 245–276.

Post, T., M. J. Van den Assem, G. Baltussen, and R. H. Thaler, "Deal or no deal? Decision making under risk in a large-payoff game show," *The American Economic Review*, 2008, *98* (1), 38–71.

Roberts, Mark J. and Dylan Supina, "Output price, markups, and producer size," *European Economic Review*, 1996, *40* (3), 909–921.

_ **and** _ , "Output price and markup dispersion in micro data: The roles of producer heterogeneity and noise," *Advances in Applied Microeconomics*, 2001, *9*, 1–36.

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

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

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

Syverson, C., "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 2004, *112* (6), 1181–1222.

_ , "What Determines Productivity?," *Journal of Economic Literature*, 2011, *49* (2).

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

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

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

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

Weyl, E. G. and M. Fabinger, "Pass-through as an Economic Tool," *University of Chicago, mimeo*, September 2012.

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

A Appendix: Proofs

A.1 A Folk Theorem

In this context, we need to define the policy space. Provided M_e and $q(c)$, and assuming without loss of generality that all of $q(c)$ is consumed, allocations are determined. The only question remaining is what class of $q(c)$ the policymaker is allowed to choose from. A sufficiently rich class for our purposes is $q(c)$ which are positive and continuously differentiable on some closed interval and zero otherwise. This follows from the basic principle that a policymaker will utilize low cost firms before higher cost firms. Proposition 1 shows without loss of generality that quantities produced at the market equilibrium are continuously differentiable. So 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 use the following shorthand throughout the proofs: $G(x) \equiv \int_0^x g(c)dc$, $R(x) \equiv \int_0^x c^{\rho/(\rho-1)} g(c)dc$.

Proof of Proposition 1. Our assumptions imply the existence of a unique market equilibrium, which guarantees $R(c)$ is finite for admissible c.First note that at both the market equilibrium and the social optimum, $L/M_e = f_e + fG(c_d)$ implies utility of zero so in both cases $L/M_e > f_e + fG(c_d)$. The policymaker's problem is

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

where the maximum is taken over choices of M_e, c_d, $q \in \mathscr{Q}_{[0,c_d]}$. We will exhibit a globally optimal $q^*(c)$ for each fixed (M_e, c_d) pair, reducing the policymaker's 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 $\forall\lambda$.[19] Now for fixed (M_e, c_d), consider the problem of finding q^* given by

$$\max_{q\in\mathscr{Q}_{[0,c_d]}} V(q) \text{ subject to } H(q) = L/M_e - f_e - fG(c_d). \tag{2}$$

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

Following Troutman (1996), if some q^* maximizes $V(q) - \lambda H(q)$ on $\mathscr{Q}_{[0,c_d]}$ for some λ and satisfies the constraint then it is a solution to Equation (2). For any λ, a sufficient condition for some q^* to be a global maximum on $\mathscr{Q}_{[0,c_d]}$ is

$$D_2 v(c, q^*(c)) \;\; = \;\; \lambda D_2 h(c, q^*(c)). \tag{3}$$

This follows because (3) implies for any such q^*, $\forall \xi$ s.t. $q^* + \xi \in \mathscr{Q}_{[0,c_d]}$ 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. Condition (3) is $\rho q^*(c)^{\rho-1} g(c) = \lambda c g(c)$ which implies $q^*(c) = (\lambda c/\rho)^{1/(\rho-1)}$.[20] 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 an appropriate λ since for fixed λ we have

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

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

$$q^*(c) \;\; = \;\; c^{1/(\rho-1)} \left(L/M_e - f_e - fG(c_d)\right) / LR(c_d) \tag{4}$$

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

$$W(M_e, c_d) \equiv M_e L \int_0^{c_d} q^*(c)^\rho g(c) dc = M_e L^{1-\rho} [L/M_e - f_e - fG(c_d)]^\rho R(c_d)^{1-\rho}. \tag{5}$$

Direct investigation yields a unique solution to the FOC of $M_e^*(c_d) = (1-\rho)L/(f_e + fG(c_d))$ and $d^2W/d^2M_e < 0$ so this solution maximizes W.
Finding c_d. Finally, we have maximal welfare for each fixed c_d from Equation (5), 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

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

Defining $B(c_d) \equiv \ln R(c_d) - \ln\left(f_e + fG(c_d)\right)$ 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 rear-

[20]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.

ranging using $R'(c_d) = c_d^{\rho/(\rho-1)} g(c_d)$ yields

$$B'(c_d) = \left\{ c_d^{\rho/(\rho-1)} - R(c_d) f / [f_e + f G(c_d)] \right\} / g(c_d) R(c_d). \tag{6}$$

Since $\lim_{c_d \longrightarrow 0} c_d^{\rho/(\rho-1)} = \infty$ and $\lim_{c_d \longrightarrow \infty} c_d^{\rho/(\rho-1)} = 0$ while $R(c_d)$ and $G(c_d)$ are bounded, there is a positive interval $[a,b]$ outside of which $B'(x) > 0$ for $x \leq a$ and $B'(x) < 0$ for $x \geq b$. Clearly $\sup_{x \in (0,a]} B(x), \sup_{x \in [b,\infty)} B(x) < \sup_{x \in [a,b]} B(x)$ and therefore any global maximum of B occurs in (a,b). Since B is continuously differentiable, a maximum exists in $[a,b]$ and all maxima occur at critical points of B. From Equation (6), $B'(c_d) = 0$ iff $R(c_d)/c_d^{\rho/(\rho-1)} - G(c_d) = f_e/f$. For c_d that satisfy $B'(c_d) = 0$, M_e^* and q^* are determined and inspection shows the entire system corresponds to the market allocation. Therefore B has a unique critical point, which is a global maximum that maximizes welfare.

A.2 VES Market Allocation

Proof of Lemma 1. Define feasible aggregate real revenues for $(M_e, c_d, q(c))$ that satisfy the resource constraint by $\mathscr{R}(M_e, c_d, q(c)) \equiv M_e \int_0^{c_d} L u'(q(c))\, q(c)\, dG$. First note that $\mathscr{R}$ is bounded. This follows from Assumption 3 by the following argument. Fix $\left(\widehat{M}_e, \widehat{c}_d\right)$ and define $\widehat{q}(c)$ by

$$u''(\widehat{q}(c))\,\widehat{q}(c) + u'(\widehat{q}(c)) = u'(\widehat{q}(c))\,[1 - \mu(\widehat{q}(c))] = \widehat{\delta}\left(\widehat{M}_e, \widehat{c}_d\right) \cdot c \tag{7}$$

where $\widehat{\delta}\left(\widehat{M}_e, \widehat{c}_d\right)$ is the infinimum of such $\widehat{\delta}$ that satisfy the resource constraint (so that it exactly holds). By Assumption 1.2 there is an $m \in (0,1)$ such that (for the analogue $\widehat{\pi}(c)$ of the profit expression for the market)

$$L\left(u'(\widehat{q}(c))/\widehat{\delta}\right)\widehat{q}(c) = \widehat{\pi}(c) + f + Lc\widehat{q}(c) \leq \frac{1}{m} Lc\widehat{q}(c).$$

Therefore aggregate real revenues at $\widehat{q}(c)$ are bounded as:

$$\begin{aligned}
\mathscr{R}\left(\widehat{M}_e,\widehat{c}_d,\widehat{q}(c)\right) &= \widehat{\delta}\widehat{M}_e\int_0^{\widehat{c}_d} L\left(u'\left(\widehat{q}(c)\right)/\widehat{\delta}\right)\widehat{q}(c)\,dG \\
&\leq \widehat{\delta}\frac{\widehat{M}_e}{m}\int_0^{\widehat{c}_d} Lc\widehat{q}(c)\,dG \\
&= \widehat{\delta}\frac{\widehat{M}_e}{m}\int_0^{\widehat{c}_d} Lc\left(u'\right)^{-1}\left(\frac{\widehat{\delta}c}{1-\mu\left(\widehat{q}(c)\right)}\right)dG \\
&\leq \widehat{\delta}\frac{1}{m}\frac{L^2}{f_e}\int_0^{\infty} c\left(u'\right)^{-1}\left(\frac{\widehat{\delta}c}{1-m}\right)dG \\
&= \delta^{\text{finite}}\frac{1-m}{m}\frac{L^2}{f_e}\int_0^{\infty} c\left(u'\right)^{-1}\left(\delta^{\text{finite}}c\right)dG \qquad (8)
\end{aligned}$$

where the second to last line follows from Assumption 1.2 and $\widehat{M}_e \leq L/f_e$ (from the resource constraint), while the last by a change of variable in the integrand. Assumption 3.2 implies the last line, 8, is finite.

We will now show that $\mathscr{R}\left(\widehat{M}_e,\widehat{c}_d,\cdot\right)$ is bounded for quantity distributions that satisfy the resource constraint. Let $q(c)$ satisfy the resource constraint at $\left(\widehat{M}_e,\widehat{c}_d\right)$, so that since $\int_0^{\widehat{c}_d} Lcq(c)\,dG < \infty$, it follows that $G(\{c:\ cq(c)=\infty\})=0$. Since G is absolutely continuous, $G(\{c:\ c=0\})=0$ and therefore $G(\{c:\ q(c)=\infty\}) \leq G(\{c:\ cq(c)=\infty\})+G(\{c:\ c=0\})=0$, i.e. $q(c)$ is bounded almost everywhere and similarly $\widehat{q}(c)$. Thus, for almost every c, there is a $\gamma(c)\in[0,1]$ such that for $q_{\xi}(c)\equiv \xi q(c)+(1-\xi)\widehat{q}(c)$, by the mean value theorem

$$u'(q(c))\,q(c)-u'(\widehat{q}(c))\,\widehat{q}(c) = u'\left(q_{\gamma(c)}\right)\left(1-\mu\left(q_{\gamma(c)}\right)\right)\left(q(c)-\widehat{q}(c)\right).$$

Appealing to this identity, we have for indicator function $1_A(x) = 1$ if $x \in A$ and 0 otherwise,

$$\begin{aligned}
\mathscr{R}\left(\widehat{M}_e, \widehat{c}_d, q(c)\right) - \mathscr{R}\left(\widehat{M}_e, \widehat{c}_d, \widehat{q}(c)\right) &= \widehat{M}_e L \int_0^{\widehat{c}_d} \left[u'(q(c))\, q(c) - u'(\widehat{q}(c))\, \widehat{q}(c)\right] dG \\
&= \widehat{M}_e L \int_0^{\widehat{c}_d} u'\left(q_{\gamma(c)}\right)\left(1 - \mu\left(q_{\gamma(c)}\right)\right)\left(q(c) - \widehat{q}(c)\right) dG \\
&\leq \frac{L^2}{f_e} \int_0^{\widehat{c}_d} u'\left(q_{\gamma(c)}\right)(1-m)\left(q(c) - \widehat{q}(c)\right) dG \\
&\leq \frac{L^2}{f_e} \int_0^{\widehat{c}_d} u'\left(q_{\gamma(c)}\right)(1-m)\left(q(c) - \widehat{q}(c)\right) 1_{q(c) > \widehat{q}(c)}(c)\, dG \\
&\leq \frac{L^2}{f_e} \int_0^{\widehat{c}_d} u'(\widehat{q}(c))(1-m)\left(q(c) - \widehat{q}(c)\right) 1_{q(c) > \widehat{q}(c)}(c)\, dG \\
&= \frac{L^2}{f_e} \int_0^{\widehat{c}_d} \frac{\widehat{\delta} c (1-m)}{1 - \mu(\widehat{q}(c))}\left(q(c) - \widehat{q}(c)\right) 1_{q(c) > \widehat{q}(c)}(c)\, dG \\
&= \widehat{\delta} \frac{L^2}{f_e} \int_0^{\widehat{c}_d} \frac{c(1-m)}{m}\left(q(c) - \widehat{q}(c)\right) 1_{q(c) > \widehat{q}(c)}(c)\, dG \\
&\leq \widehat{\delta} \frac{L^2}{f_e} \int_0^{\widehat{c}_d} \frac{(1-m)}{m} c q(c)\, dG
\end{aligned}$$

where the third line follows from $\widehat{M}_e \leq L/f_e$ and $\mu \leq m$, the fifth from $u'' < 0$, the sixth from $u'(\widehat{q}(c)) = \widehat{\delta} c / (1 - \mu(\widehat{q}(c)))$ and the seventh from $\mu \geq 1 - m$. Finally the last term is finite because $q(c)$ satisfies the resource constraint.

Since for any fixed pair $\left(\widehat{M}_e, \widehat{c}_d\right)$ we know $\mathscr{R}\left(\widehat{M}_e, \widehat{c}_d, \cdot\right)$ is bounded by, say $B\left(\widehat{M}_e, \widehat{c}_d\right)$, the positive and convex problem of minimizing $B\left(\widehat{M}_e, \widehat{c}_d\right) - \mathscr{R}\left(\widehat{M}_e, \widehat{c}_d, \cdot\right)$ then satisfies the sufficient conditions for an optimum exactly at $\widehat{q}(c)$ with constraint multiplier $\widehat{\delta}\left(\widehat{M}_e, \widehat{c}_d\right)$. Consequently, Equation (8) is an upper bound for $\mathscr{R}$. This also implies the maximal aggregate revenue possible may be found by evaluating

$$R\left(\widehat{M}_e, \widehat{c}_d\right) \equiv \mathscr{R}\left(\widehat{M}_e, \widehat{c}_d, \widehat{q}(c)\right) + \widehat{\delta}\left(\widehat{M}_e, \widehat{c}_d\right)\left[L - \widehat{M}_e\left(\int_0^{\widehat{c}_d} \left(Lc\widehat{q}(c) + f\right) dG(c) + f_e\right)\right]$$

on $\mathscr{D} \equiv [0, L/f_e] \times [0, \infty)$, on which R is continuously differentiable and since $G^{-1}([0,\infty)) = [0,1]$, has at least one maximum $\left(M_e^*, c_d^*\right)$, which by above is finite. Clearly $M_e^* \in \{0, L/f_e\}$ implies zero revenues, and trivially there is some $\underline{R} > 0$ for which $R\left(M_e^*, c_d^*\right) \geq \underline{R}$ so $M_e^* \in (0, L/f_e)$. The envelope theorem applied to dR/dM_e at M_e^* then implies (with $*$ denoting the revel ant maximizing terms):

$$\delta^*\left(M_e^*, c_d^*\right) = \int_0^{c_d^*} u'\left(q^*\left(c\right)\right) q^*\left(c\right) dG / \left(\int_0^{c_d^*} \left[cq^*\left(c\right) + \frac{f}{L}\right] dG + \frac{f_e}{L}\right) = \frac{R\left(M_e^*, c_d^*\right)}{L} \tag{9}$$

and thus $\delta^* \geq \underline{R}/L$. Now consider that $c_d^* = 0$ implies zero revenues so $c_d^* > 0$ and at any positive, finite $\widehat{c}_d$:

$$dR/dc_d = \widehat{M}_e L\left[u'\left(\widehat{q}\left(\widehat{c}_d\right)\right)\widehat{q}\left(\widehat{c}_d\right) - \widehat{\delta}\left(\widehat{c}_d\widehat{q}\left(\widehat{c}_d\right) + f/L\right)\right] = \widehat{M}_e L\widehat{\delta}\left[\frac{\mu\left(\widehat{c}_d\right)}{1-\mu\left(\widehat{c}_d\right)}\widehat{c}_d\widehat{q}\left(\widehat{c}_d\right) - f/L\right].$$

It follows that if $\widehat{c}_d\widehat{q}\left(\widehat{c}_d\right) < m/\left(1-m\right)\cdot f/L$, then $dR/dc_d < 0$. In particular,

$$\widehat{c}_d\widehat{q}\left(\widehat{c}_d\right) = u'\left(\widehat{q}\left(\widehat{c}_d\right)\right)\widehat{q}\left(\widehat{c}_d\right)\left(1-\mu\left(\widehat{q}\left(\widehat{c}_d\right)\right)\right)/\widehat{\delta} \leq u'\left(\widehat{q}\left(\widehat{c}_d\right)\right)\widehat{q}\left(\widehat{c}_d\right)\left(1-m\right)/\left(\underline{R}/L\right). \tag{10}$$

Since $\widehat{\delta}$ is increasing in $\widehat{c}_d$ (from the resource constraint), $\widehat{\delta}\widehat{c}_d$ is increasing in $\widehat{c}_d$ and goes to infinity as $\widehat{c}_d$ does, which implies $\widehat{q}\left(\widehat{c}_d\right)$ goes to zero as $\widehat{c}_d$ goes to infinity. Since $\lim_{q\longrightarrow 0} u'\left(q\right)q = 0$, this implies $\widehat{c}_d\widehat{q}\left(\widehat{c}_d\right) \longrightarrow 0$ as $\widehat{c}_d \longrightarrow 0$ through Equation (10). Therefore for all sufficiently large $\widehat{c}_d$, $dR/dc_d < 0$ so it must be that c_d^* is finite. By the envelope theorem, at c_d^*, it must be that

$$u'\left(q^*\left(c_d^*\right)\right)q^*\left(c_d^*\right) = \delta^*\left(c_d^* q^*\left(c_d^*\right) + f/L\right). \tag{11}$$

Now consider the equations that uniquely fix $\left(M_e, c_d, q\left(c\right)\right)$ in the market allocation, namely:

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

The first line is Equation (11), while the second is Equation (9), the third is Equation (7) and the fourth holds from the definition of δ^*. Since these conditions completely characterize every market equilibrium, the uniqueness of the market equilibrium guarantees a unique solution here as well.

A.3 Static Distortion Results

Proof of Proposition 4. The result relies on the following relationship we first prove:

$$\overline{\sigma} \equiv \sup_{c \leq c_d^{\text{mkt}}} \varepsilon\left(q^{\text{mkt}}(c)\right) > \delta/\lambda > \inf_{c \leq c_d^{\text{opt}}} \varepsilon\left(q^{\text{opt}}(c)\right) \equiv \underline{\sigma}. \tag{12}$$

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

$$\delta/\overline{\sigma} = M_e^{\text{mkt}} \int_0^{c_d^{\text{mkt}}} \left(\varepsilon\left(q^{\text{mkt}}(c)\right)/\overline{\sigma}\right) u\left(q^{\text{mkt}}(c)\right) dG < M_e^{\text{mkt}} \int_0^{c_d^{\text{mkt}}} u\left(q^{\text{mkt}}(c)\right) dG \tag{13}$$

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 (12). Now note that

$$\left[u''\left(q^{\text{mkt}}(c)\right) q^{\text{mkt}}(c) + u'\left(q^{\text{mkt}}(c)\right)\right]/\delta = c, \qquad u'\left(q^{\text{opt}}(c)\right)/\lambda = c. \tag{14}$$

And it follows from Equations (14) we have

$$\left[1 - \mu\left(q^{\text{mkt}}(c)\right)\right] \cdot u'\left(q^{\text{mkt}}(c)\right)/u'\left(q^{\text{opt}}(c)\right) = \delta/\lambda. \tag{15}$$

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 (15) 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 (15), 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 $c \neq c^*$.

Proof of Proposition 5. 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:

$$\max_{M_e, c_d, q(c)} L M_e \int_0^{c_d} v_\alpha(q(c)) dG \text{ subject to } L \geq M_e \left(\int_0^{c_d} Lcq(c) + f dG + f_e\right). \tag{16}$$

Let the Lagrange multiplier associated with each α in Equation (16) be written as $\beta(\alpha)$. By appealing to the envelope theorem and differentiating (16) 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 show this is equivalent to

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

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 $\varepsilon' \lesseqgtr 0$. Note that Equation (16) 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 $\varepsilon' > 0$.

A.4 Welfare Gains from Trade

The sufficient condition for gains from trade follows from differentiating $U = M_e \int u(q)dG = \delta/\bar{\varepsilon}$ where the average elasticity of utility is $\bar{\varepsilon} \equiv \int \varepsilon u dG / \int u dG$. An increase in market size raises the marginal utility of income at the rate of average markups $d\ln\delta/d\ln L = \int \mu pq dG / \int pq dG \equiv \tilde{\mu}$. From $d\ln\delta/d\ln L$ and $d\ln\bar{\varepsilon}/d\ln L$, the change in welfare is

$$\frac{d\ln U}{d\ln L} = \tilde{\mu}\left[1 + \int \frac{1-\mu-\bar{\varepsilon}}{\mu+\mu' q/(1-\mu)} \frac{\varepsilon u}{\bar{\varepsilon}\int u dG} dG\right] + \left[\frac{u_d}{\int u dG} \frac{c_d g(c_d)}{\bar{\varepsilon}(1-\mu_d)} (\varepsilon_d - \bar{\varepsilon})(\tilde{\mu} - \mu_d)\right].$$

When preferences are aligned, the second term in square brackets is positive because μ and $(1-\varepsilon)$ move in the same direction. The first term in square brackets is also positive when preferences are aligned, given the regularity condition that $(\mu + \mu' q/(1-\mu))' \leq 0$.

Proof of Proposition 6. Following the discussion above, it is sufficient to show that for $\gamma \equiv \varepsilon(\mu + \mu' q/(1-\mu))^{-1}$,

$$1 + \int \frac{1-\mu-\bar{\varepsilon}}{\mu+\mu' q/(1-\mu)} \frac{\varepsilon u}{\bar{\varepsilon}\int u dG} dG = \int \left[1 - \bar{\varepsilon} + \mu' q/(1-\mu)\right] \frac{\gamma u}{\bar{\varepsilon}\int u dG} dG \geq 0. \tag{17}$$

This clearly holds for $\mu' \geq 0$, and for the other case where preferences are aligned, we have $\mu' < 0 < \varepsilon'$. Expanding Equation (17) for $\bar{\gamma} \equiv \int \gamma \cdot (u/\int u dG) dG$ shows that

$$\int \left[1 - \bar{\varepsilon} + \mu' q/(1-\mu)\right] \frac{\gamma u}{\bar{\varepsilon}\int u dG} dG = [1 - \bar{\varepsilon} - \overline{\mu}]\,\bar{\gamma}/\bar{\varepsilon} + 1 + \int [\overline{\mu} - \mu] \frac{\gamma u}{\bar{\varepsilon}\int u dG} dG.$$

Since $\varepsilon' > 0$, $1 - \varepsilon - \mu > 0$ and $[1 - \bar{\varepsilon} - \overline{\mu}]\,\bar{\gamma}/\bar{\varepsilon} > 0$. Therefore, it is sufficient to show that $1 + \int [\overline{\mu} - \mu] \frac{\gamma u}{\bar{\varepsilon}\int u dG} dG > 0$. This sufficient condition is equivalent to

$$\int (\mu + \varepsilon) \frac{u}{\int u dG} dG \geq \int \mu \frac{\gamma u}{\bar{\gamma}\int u dG} dG \tag{18}$$

Since $\int \gamma \cdot (u/\bar{\gamma}\int u dG) dG = 1$ and $d\mu/dc > 0$, it follows that if $d\gamma/dc < 0$, then Equation (18) holds by stochastic dominance. As $d\gamma/dc < 0$ iff $d\gamma/dq > 0$, and $\varepsilon' \geq 0$, it is sufficient that $\mu + \mu' q/(1-\mu)$ is decreasing in q , which is true by assumption.