SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS*

Monika Mrázová†
University of Geneva,
CEPR and CESifo

J. Peter Neary‡
University of Oxford,
CEPR and CESifo

Mathieu Parenti§
ULB and CEPR

November 20, 2020

*This paper was first presented at ETSG 2014 in Munich under the title "Technology, Demand, and the Size Distribution of Firms". We are particularly grateful to Jan De Loecker and Julien Martin for assisting us with the data, to Stéphane Guerrier for computational advice, to our editor, Aviv Nevo, to five referees, to our conference discussants, Costas Arkolakis, Luca Macedoni, Marc Melitz, Gianmarco Ottaviano, Ina Simonovska, Frank Verboven, and Tianhao Wu, and also to Abi Adams, Andy Bernard, Bastien Chopard, Jonathan Dingel, Peter Egger, Xavier Gabaix, Basile Grassi, Arshia Hashemi, Joe Hirschberg, Oleg Itskhoki, Jérémy Lucchetti, Rosa Matzkin, Isabelle Méjean, Steve Redding, Kevin Roberts, Stefan Sperlich, Jens Südekum, Gonzague Vannoorenberghe, Maria-Pia Victoria-Feser, Frank Windmeijer, and participants at various conferences and seminars, for helpful comments and discussions. Monika Mrázová thanks the Fondation de Famille Sandoz for funding under the "Sandoz Family Foundation – Monique de Meuron" Programme for Academic Promotion. Peter Neary thanks the European Research Council for funding under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC grant agreement no. 295669.

†Geneva School of Economics and Management (GSEM), University of Geneva, Bd. du Pont d'Arve 40, 1211 Geneva 4, Switzerland; monika.mrazova@unige.ch.

‡Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK; peter.neary@economics.ox.ac.uk.

§ECARES, Université Libre de Bruxelles, Av. F.D. Roosevelt, 42, 1050 Brussels, Belgium; mathieu.parenti@ulb.ac.be.

Abstract

We characterize the relationship between the distributions of two variables linked by a structural model. We then show that, in models of heterogeneous firms in monopolistic competition, this relationship implies a new demand function that we call "CREMR" (Constant Revenue Elasticity of Marginal Revenue). This demand function is the only one that is consistent with productivity and sales distributions having the same form (whether Pareto, lognormal, or Fréchet) in the cross section, and it is necessary and sufficient for Gibrat's Law to hold over time. Among the applications we consider, we use our methodology to characterize misallocation across firms; we derive the distribution of markups implied by any assumptions on demand and productivity; and we show empirically that CREMR-based markup distributions provide an excellent parsimonious fit to Indian firm-level data, which in turn allows us to calculate the proportion of firms that are of sub-optimal size in the market equilibrium.

(150 words)

Keywords: CREMR Demands; Gibrat's Law; Heterogeneous Firms; Lognormal versus Pareto Distributions; Quantifying Misallocation; Sales and Markup Distributions.

JEL Classification: F15, F23, F12

1 Introduction

The hypothesis of a representative agent has provided a useful starting point in many fields of economics. However, sooner or later, both intellectual curiosity and the exigencies of matching empirical evidence make it desirable to take account of agent heterogeneity. In many cases, this involves constructing models with three components. First is a distribution of agent characteristics, usually assumed exogenous; second is a model of individual agent behavior; and third, implied by the first two, is a predicted distribution of outcomes. Models of this kind are now pervasive in many fields, including income distribution, optimal income taxation, macroeconomics, and urban economics.[1] In the field of international trade they have rapidly become the dominant paradigm, since the increasing availability of firm-level export data from the mid-1990s onwards undermined the credibility of representative-firm models, and stimulated new theoretical developments. A key contribution was Melitz (2003), who built on Hopenhayn (1992) to derive an equilibrium model of monopolistic competition with heterogeneous firms. In this setting, the model structure combines assumptions about the distribution of firm productivity and about the form of demand that firms face, and from these derives predictions about the distribution of firm sales. Such models have provided a fertile laboratory for studying a wide range of problems relating to the process of globalization. However, with a few exceptions to be discussed below, we know little about how different assumptions about the distributions of two variables and the structural model that links them are related to each other.

In this paper we first provide a complete characterization of this problem in the general case. This reveals how a structural model constrains the choice of assumptions and the outcomes that are consistent with them. We then show that, in models of heterogeneous firms in monopolistic competition, this implies a new demand function that we call "CREMR" (Constant Revenue Elasticity of Marginal Revenue). This demand function is the only one

[1] For examples, see Stiglitz (1969), Mirrlees (1971), Krusell and Smith (1998), and Behrens et al. (2014), respectively.

that is consistent with productivity and sales distributions having the same form (whether Pareto, lognormal, or Fréchet) in the cross section; and, with additive separability, it is necessary and sufficient for Gibrat's Law, a central result in the dynamics of firm size and industry structure, which predicts that the growth rate of firm sales is independent of firm size. Among the applications we consider, we use our methodology to characterize misallocation across firms; we derive the distribution of markups implied by any assumptions on demand and productivity; and we show empirically that CREMR-based markup distributions provide an excellent parsimonious fit to Indian firm-level data, which in turn allows us to calculate the proportion of firms that are of sub-optimal size in the market equilibrium.

Existing results in the theoretical literature on heterogeneous firms highlight important special cases in models of monopolistic competition, but give little guidance as to whether the insights can be generalized. Helpman et al. (2004) and Chaney (2008) considered what can be called the canonical model in this field, where firm productivities have a Pareto distribution and demands are CES. They showed that in this case the implied distribution of sales is also Pareto. Head et al. (2014) derived a second result with a similar flavor: lognormal productivities plus CES demands imply a lognormal distribution of firm sales. Finally, the literature on Gibrat's Law has shown that the rate of growth of a firm's sales is independent of its size, following both idiosyncratic and industry-wide productivity shocks in monopolistic competition with CES demands. (See Luttmer (2007, 2011) and Arkolakis (2010a, 2010b, 2016).)

All these results give sufficient conditions for the distribution of sales or sales growth to take a particular form. This leaves open the question of whether there are necessary conditions that can be stated, and in particular whether any demand functions other than CES are consistent with results of this kind. CES demands have great analytic convenience: their tractability has made it possible to extend CES-based models to incorporate various real-world features of the global economy, such as outsourcing, multi-product firms, and

global value chains.[2] However, in a monopolistically competitive setting they also have strong counterfactual implications. In particular, they imply that markups are constant across space and time: in a cross section, all firms should have the same markup in all markets; while, in time series, exogenous shocks such as globalization cannot affect markups and so competition effects will never be observed. Trade economists have been uneasy with these stark predictions for some time, and a number of contributions has explored the implications of relaxing the CES assumption, though to date without considering their implications for sales and markup distributions.[3] Only recently has it become possible to confront the predictions of CES-based models with data, following the development of techniques for measuring markups that do not impose assumptions about market structure or the functional form of demand. In particular, De Loecker et al. (2016) show that the distribution of markups from a sample of Indian firms is very far from being concentrated at a single value. (We discuss their data in more detail in Section 6.1 below.) A possible explanation is that such markup heterogeneity arises from aggregation across sectors with different elasticities of substitution. However, Lamorgese et al. (2014), who use data on Chilean firms, show that markup heterogeneity persists when the data are disaggregated by sector. Taken together, this evidence suggests that markup distributions are far from the Dirac form implied by CES demands, but the literature to date has paid little attention to the form of these distributions implied by alternative assumptions and how well they match the data.

In addition to our substantive results, we make two technical contributions. First, we introduce the "Generalized Power Function" class of probability distributions. This nests many two-parameter distributions, including Pareto, lognormal and Fréchet, and allows compact proofs that apply to all these cases. Second, we introduce the property of "h-

[2]See Antràs and Helpman (2004), Bernard et al. (2011), and Antràs and Chor (2013), respectively.

[3]The implications of demand functions other than CES have been considered by Melitz and Ottaviano (2008), Zhelobodko et al. (2012), Fabinger and Weyl (2012), Bertoletti and Epifani (2014), Simonovska (2015), Feenstra and Weinstein (2017), Mrázová and Neary (2017), Parenti et al. (2017), Arkolakis et al. (2018), and Feenstra (2018), among others.

reflection" of two distributions: the distribution of z is a h-reflection of that of y if the distributions of y and $h(z)$ are members of the same family of distributions, where $h(z)$ is a monotonically increasing function. This provides a unifying principle for a range of new results relating the distributions of firm characteristics and the economic model that links them.

The rest of the paper proceeds as follows. Section 2 states a general proposition which characterizes the form that distributions of agent characteristics and models of agent behavior must take if they are to be mutually consistent. Section 3 applies this result in the context of heterogeneous firms in monopolistic competition to characterize the links between the distributions of firm productivity and firm sales, and the structure of demand. This Section highlights our new CREMR demand function, and explores its properties. Sections 4 and 5 apply these results to distributions of output (in both the market equilibrium and the social optimum), and markups, respectively. Section 6 provides a quantitative illustration of various theoretical results from previous sections. First, we take to data a selection of markup distributions implied by different assumptions about demand and the distribution of firm productivities. Out of the selected alternatives, CREMR demands perform the best. We then use this best-fitting specification to quantify the degree of misallocation in a novel way. Finally, Section 7 concludes, while the Appendix and Online Appendix give proofs of propositions as well as further technical details and robustness checks.

2 Characterizing Links Between Distributions

The first main result of the paper links the distributions of two agent characteristics to a general specification of the relationship between them: until Section 3 we make no assumptions about whether either characteristic is exogenous or endogenous, nor about the underlying structural model that relates them. We assume a hypothetical dataset of a continuum of agents, which reports for each agent i its characteristics $y(i)$ and $z(i)$, both of which are

monotonically increasing functions of i.[4] Formally:

Assumption 1. $\{i, y(i), z(i)\} \in \Omega \times \mathbb{R}^2_+$, *where* Ω *is the set of agents, with both* $y(i)$ *and* $z(i)$ *monotonically increasing functions of* i.

Examples of $y(i)$ and $z(i)$ in models of heterogeneous firms include productivity, sales and markups.

In addition, we assume that the distributions of the two agent characteristics share a common parametric structure:

Definition 1. *A family of probability distributions is a member of the "Generalized Power Function" ["GPF"] class of distributions if there exists a continuously differentiable function* $H(\cdot)$ *such that the cumulative distribution function of every member of the family can be written as:*

$$G(y; \boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2} y^{\theta_2}\right) \tag{1}$$

where each member of the family corresponds to a particular value of the vector $\boldsymbol{\theta} \equiv \{\theta_0, \theta_1, \theta_2\}$.

The function $H(\cdot)$ is completely general, other than exhibiting the minimal requirements of a probability distribution: $G(\underline{y}; \boldsymbol{\theta}) = 0$ and $G(\overline{y}; \boldsymbol{\theta}) = 1$, where $[\underline{y}, \overline{y}]$ is the support of G; and, to be consistent with a strictly positive density function, $G_y > 0$, $H(\cdot)$ must satisfy the restriction: $\theta_1 H' > 0$. As we show in Appendix A.1, the great convenience of the GPF class given by (1) is that it nests many of the most widely-used families of distributions in applied economics, including Pareto, lognormal, uniform, Fréchet, Gumbel, and Weibull, as well as their truncated versions.

Given Assumption 1 and Definition 1, we can now state our main result:

[4]Conditional on monotonicity, the assumption that $y(i)$ and $z(i)$ are increasing in i is without loss of generality. For example, if $y(i)$ is increasing and $z(i)$ is decreasing, Proposition 1 can easily be reformulated using the survival function of z. By contrast, the assumption that they are monotonic in i is an important restriction, though one that is satisfied by most firm characteristics in models with uni-dimensional firm heterogeneity, on which we focus here. (For models with multi-dimensional heterogeneity, see Hallak and Sivadasan (2013), Holmes and Stevens (2014), and Harrigan and Reshef (2015).) Note that we require that monotonicity hold in theoretical models only: measured firm characteristics need not be monotonically related in the data.

Proposition 1. *Assume Assumption 1 holds. Then any two of the following imply the third:*

(A) The distribution of y *is a member of the GPF class:* $G(y;\boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2} y^{\theta_2}\right)$, $G_y > 0$;

(B) The distribution of a monotonically increasing function of z, $h(z)$, $h' > 0$, *is a member of the same family of distributions as that of* y *but with different values of* θ_1 *and* θ_2*:* $F(z;\boldsymbol{\theta'}) = G\left(h(z);\boldsymbol{\theta'}\right) = H\left(\theta_0 + \frac{\theta'_1}{\theta'_2} h(z)^{\theta'_2}\right)$, $F_z > 0$;

(C) y *is a power function of* $h(z)$*:* $y = y_0 h(z)^E$;

where the parameters are related as follows:

(i) (A) and (C) imply (B) with $\theta'_1 = E\theta_1 y_0^{\theta_2}$ *and* $\theta'_2 = E\theta_2$*; similarly, (B) and (C) imply (A) with* $\theta_1 = E^{-1}\theta'_1 y_0^{-E^{-1}\theta'_2}$ *and* $\theta_2 = E^{-1}\theta'_2$.

(ii) (A) and (B) imply (C) with $y_0 = \left(\frac{\theta_2}{\theta_1}\frac{\theta'_1}{\theta'_2}\right)^{\frac{1}{\theta_2}}$ *and* $E = \frac{\theta'_2}{\theta_2}$.

Comparing the distributions of y and $h(z)$ in (A) and (B), they are members of the same family of the GPF class, except that the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\theta'}$ are different. The $h(\cdot)$ function is completely general, except that it must be monotonically increasing from the monotonicity restriction on F: $h' > 0$ since $F_z = G_y h' > 0$; and the elements of $\boldsymbol{\theta}$ can take on any values, except that θ_0 must be the same for both distributions.

Each choice of the $h(\cdot)$ function generates in turn a further family, such that the transformation $h(z)$ follows a distribution from the GPF class. Proposition 1 shows that these families are intimately linked via a simple power function that expresses one of the two agent characteristics as a transformation of the other. We say that the distribution of z is a h-reflection of that of y:[5]

Definition 2. *The distribution of* z *is a* h*-reflection of the distribution of* y *if the distributions of* y *and* $h(z)$ *are members of the same family of distributions.*

[5]The property of h-reflection is not symmetric in general: the fact that the distribution of z is a h-reflection of the distribution of y does not imply that the distribution of y is a h-reflection of the distribution of z. Also h-reflection is not in itself related to the GPF class, though all the cases we consider in the paper assume that the distributions of both y and $h(z)$ are members of a family of the GPF class.

In the remainder of the paper, we apply Proposition 1 to the setting of heterogeneous firms in monopolistic competition. Our theoretical results can be categorized by the type of h-reflection they exhibit. One central case is where $h(z)$ is the identity transformation, $h(z) = z$. We call this case "self-reflection", since it implies from Proposition 1 that the distributions of y and z are members of the same family. This case proves particularly useful when we consider distributions of firm sales and the rate of growth of firm sales in Section 3.

When we come to consider the distributions of output in the market equilibrium and the social optimum in Section 4, we will see that they exhibit "marginal-revenue reflection" and "marginal-utility reflection" of the distribution of productivity respectively. Finally, when we come to consider the distributions of sales and firm markups for a range of demand functions in Section 5, we will see that they exhibit a wide range of forms for the h function. One important case is the odds transformation, $h(z) = \frac{z}{1-z}$, where $0 \leq z \leq 1$. When the distributions of y and of an odds transformation of z are members of the same family, we say that the distribution of z exhibits "odds reflection" of that of y. This case proves particularly useful when we consider distributions of firm markups.

3 CREMR Demands

In this section we explore some implications of Proposition 1 in models of monopolistic competition with heterogeneous firms and general demands. In particular, we ask what demand functions are consistent with the distributions of firm productivity and sales revenue exhibiting self-reflection, so the two distributions are members of the same family from the GPF class though with different parameters. (Appendix A.4 gives related results for self-reflection of productivity and output and of sales and output.) As noted in the introduction, there are only two results in the literature that relate productivity and sales distributions: Helpman et al. (2004) and Chaney (2008) showed that CES demands are sufficient to bridge the gap between two Pareto distributions; and Head et al. (2014) showed that the same holds for two

lognormal distributions. Given the abundant empirical evidence that both firm productivity and sales are either Pareto or lognormal shaped, self-reflection is a natural starting point in trying to generalize these results.[6] Our results illustrate the power of Proposition 1: it leads to a complete characterization of the conditions under which self-reflection holds. This yields a new demand function that we call "CREMR", which implies functional forms for the distribution of markups for which we find strong evidence in our empirical section. In Section 3.3, we give a further motivation for self-reflection, showing that it is central to an important substantive question: when does Gibrat's Law hold in monopolistic competition? We show that, under additive separability, self-reflection of cumulated productivity shocks and sales growth rates over time is equivalent to Gibrat's Law, so Proposition 1 implies that CREMR demands are necessary and sufficient for this Law to hold in a monopolistically competitive industry. Given these important implications of CREMR demands, it is desirable to understand their properties and to consider what preferences rationalize them: Sections 3.4 and 3.5 consider these topics respectively.

We begin in Section 3.1 by introducing the monopolistically competitive setting we will use in the remainder of the paper.

3.1 The Monopolistically Competitive Setting

Consider a model of a monopolistically competitive industry with heterogeneous firms in the tradition of Melitz (2003), extended to allow for non-CES demands. Firms differ in their productivity, φ, which is drawn from a distribution $G(\varphi)$ with support $[\underline{\varphi}, \infty)$. They incur a common fixed cost f which may be zero in the case when the demand function implies a finite upper bound for marginal revenue. Each firm produces a unique good, and chooses its output x to maximize its profits π, which equal operating profits less fixed costs:

[6] Axtell (2001) and Gabaix (2009) argue that the distribution of firm sales is plausibly close to Pareto, at least in the upper tail. However, Head et al. (2014) and Bee and Schiavo (2018) argue that it is better approximated overall by a lognormal, and Fernandes et al. (2018) find that the intensive margin of firm sales is inconsistent with a Pareto productivity distribution. We return to this issue in Online Appendix B.4 below.

$$\pi(\varphi, \lambda, \tau) = \max_x \left(\left(p(x, \lambda) - \tau\varphi^{-1} \right) x - f \right) \tag{2}$$

Here, $p(x, \lambda)$ is the inverse demand function of a representative consumer faced by all firms, which depends negatively on their output level x and on λ, a common demand parameter that is exogenous to firms but endogenous to the industry. From each firm's perspective, λ is a measure of the intensity of competition which it takes as given.[7] Finally, τ is a uniform cost shifter that is common to all firms; until Section 3.3 we set this equal to one.

Maximizing profits as in (2) leads to the first-order condition, which equates marginal revenue to marginal cost:

$$p(x, \lambda) + xp_x(x, \lambda) = \varphi^{-1} \tag{3}$$

Assuming the second-order condition $2p_x(x, \lambda) + xp_{xx}(x, \lambda) < 0$ is satisfied, (3) implies that the equilibrium output and price of each firm are functions of its productivity φ and of the demand shifter λ, where the latter is the same for all firms. In Section 3.2, we suppress λ to simplify notation.

3.2 Self-Reflection in the Cross Section: Productivity and Sales

The necessary condition for self-reflection follows immediately from Proposition 1: if the distributions of productivity φ and sales r are from the same family, which can be any member of the GPF class, then they must be related by a power function:

$$\varphi = \varphi_0 r^E \tag{4}$$

[7]The specification of demand in (2) corresponds to the generalized separability class of Pollak (1972). It allows for various preference systems including additive separability as in Zhelobodko et al. (2012), Bertoletti and Epifani (2014), and Mrázová and Neary (2017). Results in Section 3.2 take a "firm's-eye" view perspective and do not depend on the micro-foundation of demand. In Section 3.3 by contrast, we invoke additive separability when discussing general-equilibrium effects.

To infer the implications of this for demand, we use two properties of a monopolistically competitive equilibrium. First, firms equate marginal cost to marginal revenue, so from (3) $\varphi = c^{-1} = \left(\frac{\partial r}{\partial x}\right)^{-1}$. Second, all firms face the same residual demand function, so firm sales conditional on output are independent of productivity φ: $r(x) = xp(x)$ and $\frac{\partial r}{\partial x} = r'(x)$.[8] Combining these with (4) gives a simple differential equation in sales revenue:

$$(r'(x))^{-1} = \varphi_0 r(x)^E \tag{5}$$

Integrating this we find that a necessary and sufficient condition for self-reflection of productivity and sales is that the inverse demand function takes the following form:

$$p(x) = \frac{\beta}{x}(x-\gamma)^{\frac{\sigma-1}{\sigma}}, \quad 1 < \sigma < \infty,\ x > \gamma\sigma,\ \beta > 0 \tag{6}$$

Calculating marginal revenue and inverting it brings us back to (4), with the constants φ_0 and E equal to $\beta^{-\frac{\sigma}{\sigma-1}}\frac{\sigma}{\sigma-1}$ and $\frac{1}{\sigma-1}$ respectively.

We are not aware of any previous discussion of the family of inverse demand functions in (6), which express expenditure $r(x) = xp(x)$ as a power function of consumption relative to a benchmark γ. Its key property, from (5), is that the elasticity of marginal revenue with respect to total revenue is constant: $E = \frac{1}{\sigma-1}$. Hence we call it the "CREMR" family, for "Constant Revenue Elasticity of Marginal Revenue." Summarizing:

Proposition 2. *The distributions of firm productivity and firm sales revenue in models of monopolistic competition with heterogeneous firms are members of the same family of the Generalized Power Function class if and only if demands take the CREMR form* (6)*.*

CREMR demands include CES demands as a special case: when γ equals zero, (6) reduces to $p(x) = \beta x^{-\frac{1}{\sigma}}$, and the elasticity of demand is constant, equal to σ. More generally, the

[8]Our approach is consistent with marginal costs being chosen endogenously by firms, either by optimizing subject to a variable cost function, as in Zhelobodko et al. (2012), or as the outcome of investment in R&D, as in Bustos (2011). However, it is not in general consistent with oligopoly, as firms may face different residual demand functions.

elasticity of demand varies with consumption, $\varepsilon(x) \equiv -\frac{p(x)}{xp'(x)} = \frac{x-\gamma}{x-\gamma\sigma}\sigma$, though it approaches σ for large firms.[9]

It is useful to consider the implications of CREMR demands combined with Pareto and lognormal distributions of productivity. Starting with the Pareto, it follows immediately as a corollary of Proposition 1 that CREMR demands are necessary and sufficient for self-reflection in this case. We state the result formally for completeness, and because it makes explicit the links that must hold between the parameters of the two Pareto distributions and the demand function.

Corollary 1. *Given Assumption 1, any two of the following imply the third: (A) The distribution of firm productivity is Pareto:* $G_{\mathcal{P}}(\varphi) = 1 - \underline{\varphi}^k \varphi^{-k}$*; (B) The distribution of firm sales revenue is Pareto:* $F_{\mathcal{P}}(r) = 1 - \underline{r}^n r^{-n}$*; (C) The demand function belongs to the CREMR family in* (6)*; where the parameters are related as follows:*

$$\sigma = \frac{k+n}{n} \quad \Leftrightarrow \quad n = \frac{k}{\sigma - 1} \qquad \text{and} \qquad \beta = \left(\frac{k+n}{k} \frac{\underline{r}^{\frac{n}{k}}}{\underline{\varphi}} \right)^{\frac{k}{k+n}} \quad \Leftrightarrow \quad \underline{r} = \beta^{\sigma} \left(\frac{\sigma - 1}{\sigma} \underline{\varphi} \right)^{\sigma - 1} \tag{7}$$

This extends a result of Chaney (2008), who showed that $n = \frac{k}{\sigma - 1}$ with Pareto productivity and CES demands.

Turning next to the lognormal, since it is also a member of the GPF class, it follows immediately from Proposition 1 that the CREMR relationship $\varphi = \varphi_0 r^E$ is necessary and sufficient for self-reflection in the lognormal case. A complication is that, except in the CES case (when the CREMR parameter γ is zero), the value of sales revenue for the smallest firm is strictly positive, whereas the lower bound of the lognormal distribution is zero.[10] However, this is not a problem since, as we show in Corollary 3 in Appendix A.1, a truncated distribution from the GPF family is itself a member of the family. Hence we have the result

[9] Note that this contrasts with CES models under oligopolistic competition (Atkeson and Burstein (2008)) where the price-elasticity of demand is equal to σ for the smallest firms only.

[10] Since $p'(x) = -\frac{\beta}{\sigma x^2}(x-\gamma)^{-\frac{1}{\sigma}}(x - \gamma\sigma)$, the output of the smallest active firm when γ is strictly positive is greater than or equal to $\gamma\sigma$, while its sales revenue is $r(x) = \beta\left(\gamma(\sigma-1)\right)^{\frac{\sigma-1}{\sigma}} > 0$. When γ is strictly negative, sales revenue is discontinuous at $x = 0$: $\lim_{x \to 0^+} r(x) = \beta(-\gamma)^{\frac{\sigma-1}{\sigma}} > 0$, but $r(0) = 0$.

(where Φ denotes the cumulative distribution function of the standard normal distribution and T denotes the fraction of potential firms that are inactive):

Corollary 2. *Given Assumption 1, any two of the following imply the third: (A) The distribution of firm productivity is truncated lognormal:* $G_{t\mathcal{LN}}(\varphi) = \frac{\Phi((\log\varphi-\mu)/s)-T}{1-T}$*; (B) The distribution of firm sales revenue is truncated lognormal:* $F_{t\mathcal{LN}}(r) = \frac{\Phi((\log r-\mu')/s')-T}{1-T}$*; (C) The demand function belongs to the CREMR family in* (6)*; where the parameters are related as follows:*

$$\sigma = \frac{s+s'}{s} \quad \Leftrightarrow \quad s' = (\sigma-1)s \tag{8}$$

$$\beta = \frac{s+s'}{s'}\exp\left(\frac{s}{s'}\mu' - \mu\right) \quad \Leftrightarrow \quad \mu' = (\sigma-1)\left(\mu + \log\left(\frac{\sigma-1}{\sigma}\beta^{\frac{\sigma}{\sigma-1}}\right)\right) \tag{9}$$

$$\underline{r} = \left(\underline{\varphi}\exp\left(\frac{s}{s'}\mu' - \mu\right)\right)^{\frac{s'}{s}} = \left(\frac{\sigma-1}{\sigma}\beta^{\frac{\sigma}{\sigma-1}}\underline{\varphi}\right)^{\sigma-1} \tag{10}$$

$$T = \Phi((\log\underline{\varphi} - \mu)/s) = \Phi\left((\log\underline{r} - \mu')/s'\right) \tag{11}$$

Just as in the Pareto case, CREMR is the only demand function that is compatible with lognormal productivity and sales.

We will see in Section 6 how these theoretical results translate to data.

3.3 Self-Reflection over Time: Gibrat's Law

Having derived the necessary and sufficient conditions for self-reflection in the cross-section, we now turn to self-reflection over time. Specifically, we show that CREMR demands are necessary and sufficient for Gibrat's Law, or "The Law of Proportionate Effect", which asserts that the rate of growth of a firm is independent of its size. There is persuasive empirical evidence in favor of the Law in general, especially for larger and older firms; see, for example, Haltiwanger et al. (2013). A variety of mechanisms has been proposed to explain this empirical regularity.[11] Early contributions, by Gibrat (1931) himself and by Ijiri and

[11] For surveys of a large literature, see Sutton (1997) and Luttmer (2010). Gibrat's Law has also been applied to the growth rate of cities. See, for example, Eeckhout (2004). We do not pursue this application

Simon (1974), gave purely stochastic explanations. In particular, if firms are subject to i.i.d. idiosyncratic shocks, these cumulate to give an asymptotic lognormal distribution of firm size, all growing at the same rate. Later work has shown how Gibrat's Law can be derived as an implication of industry equilibrium, when firms are subject to industry-wide as well as idiosyncratic shocks. Much of this work has been carried out under perfectly competitive assumptions, focusing on learning, as in Jovanovic (1982), or differential access to credit, as in Cabral and Mata (2003). The result has also been shown to hold in models of monopolistic competition by Luttmer (2007, 2011) and Arkolakis (2010a, 2010b, 2016). However, these papers assume CES demand. Putting this differently, all models that generate Gibrat's Law to date imply that prices are either equal to or proportional to marginal costs. This raises the question whether Gibrat's Law is consistent with demand functions that allow for variable markups. The following proposition shows that this is indeed the case with CREMR demands:

Proposition 3. *In monopolistic competition with additive separability, CREMR demands are necessary and sufficient for Gibrat's Law to hold following: (i) industry-wide shocks to firm productivity; and (ii) i.i.d. or AR(1) shocks to firm productivity.*

Assume that the productivity process for firm i can be written as: $\varphi_{it} = \gamma_{it}\varphi_t$, where φ_t is an industry-wide shock, common to all firms, whereas γ_{it} is a firm-specific idiosyncratic shock. We consider each of these types of shocks in turn.

Consider first an industry-wide productivity shock, as in part (i) of Proposition 3. Intuitively, it is easy to see that CREMR demands are necessary and sufficient for such a shock to have the same proportionate effect on the sales of all firms. This outcome is equivalent to a constant elasticity of sales revenue with respect to marginal cost (which is the inverse of productivity). Since marginal cost equals marginal revenue, this in turn is equivalent to the CREMR condition for self-reflection that we have already considered, which entails a

here, but it is clear that analogous results to ours can be derived in that case. As Sutton (1997) points out, different authors have considered shocks to either sales, employment, or assets. In a monopolistically competitive setting, it is natural to assume shocks to productivity, as below.

constant elasticity of marginal revenue with respect to total revenue; though the two conditions arise in different contexts: "cross-section" comparisons across firms in the case of self-reflection, "time-series" comparisons between the pre- and post-productivity-shock equilibria in the case of Gibrat's Law. Recalling Proposition 1, this suggests that CREMR demands are necessary and sufficient for Gibrat's Law to hold following industry-wide shocks to firm productivity. We can show that this holds in general equilibrium with additive separability.

Consider a uniform improvement in the productivity of all firms that we assume is exogenous and unanticipated: $\widehat{\tau} < 0$ (where a circumflex denotes a logarithmic derivative: $\widehat{\tau} = d\log\tau$, $\tau > 0$). The growth rate of sales following such a uniform productivity shock is: $g \equiv -\frac{\widehat{r}}{\widehat{\tau}} = -\frac{\tau}{r}\frac{dr}{d\tau}$. Hence Gibrat's Law ($\frac{dg}{d\varphi} = 0$) obtains when $\widehat{r}$ is independent of φ.

We first consider the effects of the shock on each firm's price and output. Starting with the household's first-order condition under additively separable preferences, $p(x,\lambda) = \lambda^{-1}u'(x)$ where $u(\cdot)$ denotes consumer's sub-utility, totally differentiate to get the proportional change in prices:

$$\widehat{p} = -\frac{1}{\varepsilon}\widehat{x} - \widehat{\lambda} \tag{12}$$

Hence the change in sales revenue is:

$$\widehat{r} = \widehat{p} + \widehat{x} = \frac{\varepsilon - 1}{\varepsilon}\widehat{x} - \widehat{\lambda} \tag{13}$$

To solve for the proportional change in outputs we totally differentiate the firm's first-order condition, (3):

$$\widehat{x} = -\frac{\varepsilon - 1}{2 - \rho}(\widehat{\tau} + \widehat{\lambda}), \tag{14}$$

where $\rho(x) \equiv -\frac{xp''(x)}{p'(x)}$ is the convexity of the demand function. Finally, we substitute (14) into (13), to obtain the change in sales revenue in terms of the cost shock $\widehat{\tau}$ and the implied

change in the intensity of competition $\widehat{\lambda}$:

$$\widehat{r} = \underbrace{-\frac{(\varepsilon-1)^2}{\varepsilon(2-\rho)}}_{(\star)}(\widehat{\tau}+\widehat{\lambda}) - \widehat{\lambda} \tag{15}$$

The way in which the change in the intensity of competition $\widehat{\lambda}$ depends on the cost shock $\widehat{\tau}$ follows from the assumptions we make about market equilibrium: in particular, it differs between the cases of free entry and a fixed number of firms. Fortunately, these differences do not matter for our purposes, since in both cases $\widehat{\tau}$ and $\widehat{\lambda}$ are the same for all firms. It follows that a necessary and sufficient condition for Gibrat's Law in this setting is that $(\star)$ is constant across firms. This term, $\frac{(\varepsilon-1)^2}{\varepsilon(2-\rho)}$, is the elasticity of revenue with respect to productivity. It is the inverse of the elasticity of marginal revenue with respect to total revenue, which as we have seen is constant if and only if demands are CREMR, in which case it equals $\frac{1}{\sigma-1}$. (See Section 3.2, and equation (51) in Appendix A.3.) This confirms that $\frac{dg}{d\varphi}=0$, i.e., with additive separability, Gibrat's Law holds following an industry-wide productivity shock in monopolistic competition, if and only if demands are CREMR.

To prove part (ii) of Proposition 3, consider now idiosyncratic shocks to firms' productivity, which can be written as follows:

$$\gamma_{it} = \gamma_{i,t-1}e^{\epsilon_{it}} \tag{16}$$

where ϵ_{it} are identically distributed shocks with zero mean and finite variance. Equation (16) implies:

$$\log\varphi_{it} = \log\gamma_{0t} + \sum_{t'=0}^{t}\epsilon_{it'} + \log\varphi_t \tag{17}$$

We consider the case of a stationary AR(1) growth rate without drift.[12] Specifically, we

[12] As long as drifts are not firm-specific, this assumption is made without loss of generality since industry-specific drifts are captured by φ_t.

allow firm growth rates to be serially correlated:

$$\epsilon_{it} = \xi \epsilon_{i,t-1} + \nu_{it} \tag{18}$$

where $\xi < 1$ and ν_{it} is white noise with constant variance υ^2. (The special case of i.i.d. growth rates is readily obtained for $\xi = 0$ and ν_{it} i.i.d.) Then, as $t \to \infty$, and provided $\log \gamma_{0t} + \log \varphi_t$ is small relative to $\log \varphi_{it}$, the distribution of φ_{it} is asymptotically lognormal:

$$\frac{\log \varphi_{it}}{t} \sim \mathcal{N}\left(0, \frac{\upsilon^2}{1-\xi^2}\right) \tag{19}$$

More generally, growth rate shocks will cumulate to give an asymptotic lognormal distribution if they admit a $MA(\infty)$ representation with absolutely summable coefficients. (See Hayashi (2000), Chapter 6, for extensions of the central limit theorem.) The final step is to recall that productivity equals the inverse of marginal revenue:

$$\varphi_{it} = \varphi_i \gamma_{it} = c_{it}^{-1} = (r'_{it})^{-1} \tag{20}$$

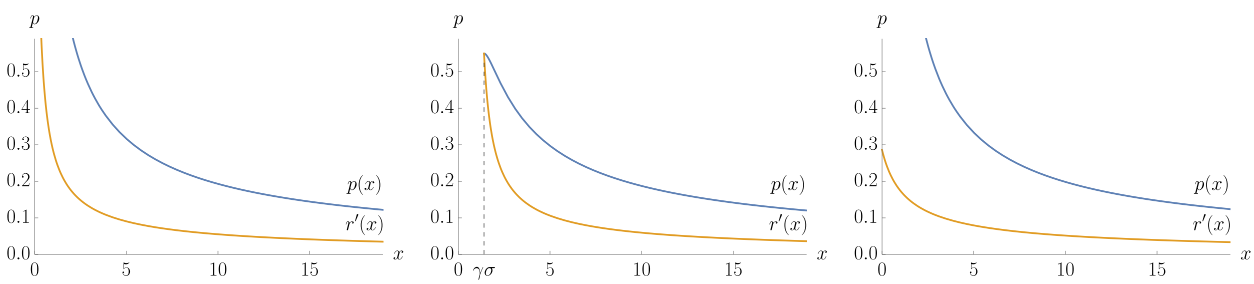
Now, we can invoke Proposition 1 and conclude that CREMR demands are necessary and sufficient for i.i.d. or AR(1) shocks to productivity to cumulate to give an asymptotic lognormal distribution of sales, with firm growth rates independent of size. Note that Proposition 1 applies in the cross-section. In the time series, idiosyncratic shocks also imply changes in λ hence the level of demand over time. As shown previously however, additive separability implies that these general equilibrium effects impact all firms proportionally, so that our cross-sectional characterization still applies.

It is also noteworthy that the above micro-foundation of Gibrat's law implies a non-stationary distribution of firm productivity and sales. Indeed, the asymptotic law expressed in (19) features a variance that increases quadratically with t. This creates a tension with the assumption that t must be large enough for the lognormal approximation to hold. This

is a well-known problem which may be solved by adding a constant term to (16), see for instance Gabaix (1999). This yields a Kesten process which leads asymptotically to a Pareto distribution in the upper tail. Proposition 1 implies in this case that CREMR is not necessary but still sufficient to obtain a Pareto distribution of sales.

Combining these results, we have proved Proposition 3: CREMR demands are necessary and sufficient for Gibrat's Law to hold in monopolistic competition with additive separability following both idiosyncratic and industry-wide shocks to firm productivity.

3.4 Properties of CREMR Demands



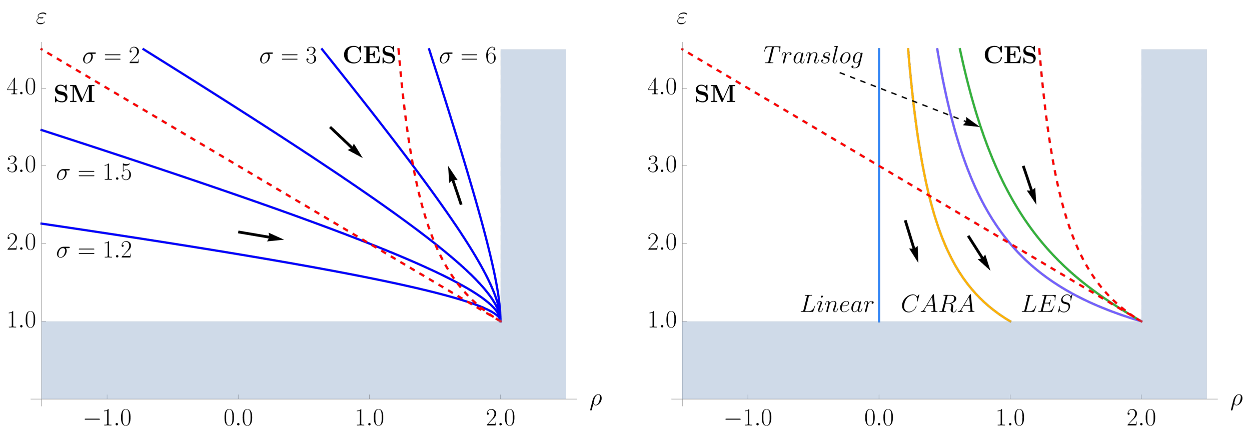
(a) $\gamma = 0$: CES (b) $\gamma > 0$: Subconvex (c) $\gamma < 0$: Superconvex

Figure 1: Examples of CREMR Demand and Marginal Revenue Functions

Note: Each panel depicts the CREMR demand and marginal revenue functions from (6) and (47) respectively, for $\beta = 1$, $\sigma = 1.4$, and γ equal to 0, 1 and -1 in panels (a), (b), and (c) respectively.

Consider next the properties of the CREMR demand function (6). They are derived formally in Appendix A.3, but can be understood by referring to the three sub-panels of Figure 1. These show three representative inverse demand curves from the CREMR family, along with their corresponding marginal revenue curves. The CES case in panel (a) combines the familiar advantage of analytic tractability with the equally familiar disadvantage of imposing strong and counter-factual properties. In particular, the markup $m \equiv \frac{p}{c}$ must be the same, equal to $\frac{\sigma}{\sigma-1}$, for all firms in all markets. By contrast, members of the CREMR family with non-zero values of γ avoid this restriction. Moreover, we show in Appendix A.3 that the sign of γ determines whether a CREMR demand function is more or less convex than a CES demand function. The case of a positive γ as in panel (b) corresponds to

demands that are "subconvex": less convex at each point than a CES demand function with the same elasticity. (See Mrázová and Neary (2019) for further discussion.) In this case the elasticity of demand falls with output, which implies that larger firms have higher markups. These properties are reversed when γ is negative as in panel (c). Now the demands are "superconvex" – more convex than a CES demand function with the same elasticity – and larger firms have smaller markups. CREMR demands thus allow for a much wider range of comparative statics responses than the CES itself.



(a) CREMR Demands

(b) Some Well-Known Demand Functions

Figure 2: Demand Manifolds for CREMR and Other Demand Functions

Note: Each curve shows the combinations of elasticity ε and convexity ρ implied by the demand function indicated. Values of ε and ρ in the shaded region are inadmissible. See text for details.

How do CREMR demands compare with other better-known demand systems? Inspecting the demand functions themselves is not so informative, as they depend on three different parameters. Instead, we use the approach of Mrázová and Neary (2017), who show that any well-behaved demand function can be represented by its "demand manifold", a smooth curve relating its elasticity $\varepsilon(x) \equiv -\frac{p(x)}{xp'(x)}$ to its convexity $\rho(x) \equiv -\frac{xp''(x)}{p'(x)}$. We show in Appendix A.3 that the CREMR demand manifold can be written in closed form as follows:

$$\rho(\varepsilon) = 2 - \frac{1}{\sigma - 1} \frac{(\varepsilon - 1)^2}{\varepsilon} \tag{21}$$

Whereas the demand function (6) depends on three parameters, the corresponding demand

manifold only depends on σ: it is invariant with respect to β and γ. Panel (a) of Figure 2 illustrates some manifolds from this family for different values of σ, while panel (b) shows the manifolds of some of the most commonly-used demand functions in applied economics: linear, CARA, Translog and LES.[13] It is clear that CREMR manifolds, and hence CREMR demand functions, behave very differently from the others. The arrows in Figure 2 denote the direction of movement as sales increase. In the empirically relevant subconvex region, where demands are less convex than the CES, CREMR demands are more concave at low levels of output (i.e., at high demand elasticities) than any of the others, which are approximately linear for small firms. As we move to larger firms, the CREMR elasticity of demand falls more slowly with convexity than any of the others. As for the largest firms, with CREMR demands they asymptote towards a demand function with elasticity equal to σ; whereas with other demand functions the largest firms either hit an upper bound of maximum profitable output (in the linear and CARA cases), or else asymptote to a Cobb-Douglas demand function with elasticity of one (in the translog and LES cases).

3.5 CREMR Preferences

Next, we ask what specifications of preferences rationalize CREMR demands. The simplest way of doing this is to assume additively separable preferences as in Section 3.3, $U = \int_{i \in X} u(x(i))\mathrm{d}i$ where X is the set of available goods. For every $i \in X$, $x(i)$ takes values in

[13] CARA demands are implied by a negative-exponential utility function, which has the same form as a constant-absolute-risk-aversion utility function in the theory of choice under uncertainty; translog demands are observationally equivalent to the almost-ideal demand system of Deaton and Muellbauer (1980); the LES or Linear Expenditure System is implied by the Stone-Geary utility function. All these manifolds, derived in Mrázová and Neary (2017), are invariant to all parameters. We confine attention to the admissible region, $\{\varepsilon > 1, \rho < 2\}$, where firms' first- and second-order conditions are satisfied. The curve labeled "*CES*" is the locus $\varepsilon = \frac{1}{\rho - 1}$, each point on which corresponds to a particular CES demand function; this is also equation (21) with $\varepsilon = \sigma$. To the right of the CES locus is the superconvex region (where demand is more convex than the CES); while to the left is the subconvex region. The curve labeled "*SM*" is the locus $\varepsilon = 3 - \rho$; to the right is the "supermodular" region (where selection effects in models of heterogeneous firms must have the conventional sign, e.g., more efficient firms serve foreign markets by foreign direct investment rather than exports); while to the left is the submodular region. (See Mrázová and Neary (2019) for further discussion.) Appendix A.3 shows that the CREMR demand manifold lies wholly in the supermodular region if and only if $\sigma \geq 2$.

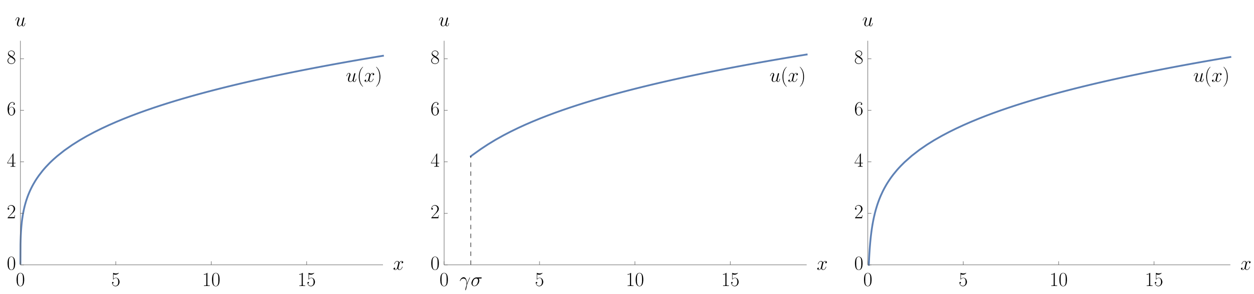
$[x_{min}, \infty)$, where x_{min} equals $\gamma\sigma$ in the subconvex case and is strictly positive but arbitrarily small in the superconvex case. This implies that $p(x(i)) = \lambda^{-1} u'(x(i))$, where λ is the marginal utility of income. Substituting for $p(x(i))$ from the CREMR demand function (6), with the demand shifter rewritten as $\beta = \lambda^{-1}\tilde{\beta}$, and integrating yields an explicit form for the sub-utility function $u(x(i))$:

$$u(x(i)) = \kappa + \tilde{\beta}\frac{\sigma}{\sigma-1}\frac{(x(i)-\gamma)^{\frac{\sigma-1}{\sigma}}}{x(i)}\left(x(i) + \gamma(\sigma-1)\, {}_2F_1\left(1, 1, 1+\frac{1}{\sigma}, \frac{\gamma}{x(i)}\right)\right) \tag{22}$$

This equals a constant of integration κ plus a primitive preference parameter $\tilde{\beta}$ times the product of two functions, one an augmented CES, the other an augmented hypergeometric:

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n}\frac{z^n}{n!}, \qquad |z| < 1, \qquad (q)_n = \frac{\Gamma(q+n)}{\Gamma(q)} \tag{23}$$

where $(q)_n$ is the (rising) Pochhammer symbol, and $\Gamma(q)$ is the gamma function. The only demand parameter that varies with income and other prices is β; it depends on λ, whose value can be recovered in a standard way.[14]



(a) $\gamma = 0$: CES

(b) $\gamma > 0$

(c) $\gamma < 0$

Figure 3: Examples of CREMR Sub-Utility Functions

Note: Each panel shows the values of sub-utility u implied by (22) as a function of x for the same parameter values as in Figure 1: $\beta = 1$, $\sigma = 1.4$, and γ equal to 0, 1 and -1 in panels (a), (b), and (c) respectively.

When γ is zero, the hypergeometric function also equals zero, and so (22) reduces to

[14] Inverting (6) yields the direct demand functions: $x(i) = (u')^{-1}(\lambda p(i))$, which can be combined with the budget constraint to obtain: $\int_{i \in X} p(i)\,(u')^{-1}(\lambda p(i))\,\mathrm{d}i = I$ (where I denotes consumer income). Solving this gives λ as a function of prices and income. Note that $x(i)$ cannot be written in closed form, but the marginal utility function is invertible provided the elasticity of demand is positive, i.e., provided $x(i) \in [x_{min}, \infty)$.

the CES utility function, $u(x(i)) = \tilde{\beta}\frac{\sigma}{\sigma-1}\,x(i)^{\frac{\sigma-1}{\sigma}} + \kappa$. Figure 3 illustrates three sub-utility functions from the CREMR family, each as a function of σ and x, for different values of γ. Panel (a) is the CES case, showing that utility is increasing and concave in x. The subconvex case in Panel (b) and the superconvex case in Panel (c) (with positive and negative values of γ respectively) deviate from the CES case in ways that parallel the ways that the corresponding demand functions differ from CES demands in Figure 1. In particular, utility is defined on the same range as the demand function. If needed, they can both be extended in an appropriate way on the entire positive range to guarantee love for variety.

In some applications it may be desirable to have a homothetic specification of preferences consistent with CREMR demands. This is not possible with additive separability (which implies homotheticity only in the CES case), but it can be done by embedding CREMR demands in the implicitly additive preferences of Kimball (1995).[15] Here the sub-functions corresponding to each good depend on the consumption of that good scaled by total utility U: $\displaystyle\int_{i\in X} \Upsilon\left(\frac{x(i)}{U}\right)\, \mathrm{d}i = 1$. Proceeding as in the additively separable case, we can combine the first-order condition $\Upsilon'\left(\frac{x(i)}{U}\right) = \lambda p(i)$ with the CREMR demand function (6) and integrate, which shows that the Υ sub-function takes the same form as $u(x(i))$ in (22).[16] Unlike the more familiar Klenow and Willis (2016) special case of Kimball preferences, the Kimball-CREMR direct utility and demand functions cannot be written in closed form, but they can still be used as a foundation for quantitative analysis of normative issues.

[15] Fally (2018) shows that CREMR demands can be integrated to give utility functions from other members of the single-aggregate Pollak (1972) generalized separability class, but only in the superconvex case.

[16] Now the direct demand functions depend on two aggregates rather than one, the true price index P and the shadow price of the budget constraint λ: $x(i) = (\Upsilon')^{-1}(\lambda p(i))\,\frac{I}{P}$. To solve for these we use two equations: first, the equation given in the text that implicitly defines U, evaluated at the optimal quantities: $\int_{i\in X}\Upsilon\left((\Upsilon')^{-1}(\lambda p(i))\right)\, \mathrm{d}i = 1$; and, second, the definition of the price index: $P = \int_{i\in X} p(i)(\Upsilon')^{-1}(\lambda p(i))\, \mathrm{d}i$. See Matsuyama and Ushchev (2017) for further details.

4 Misallocation Across Firms

Section 3 used part (ii) of Proposition 1 to back out the demands implied by assumed distributions of two firm characteristics. In this section and the next we show how part (i) of the Proposition can be used to derive distributions of firm characteristics given the distribution of productivity and the form of the demand function. In this section we show how to compare the distributions of output across firms in the market equilibrium and in the social optimum. Previous comparisons between the allocation of resources in a monopolistically competitive market and in the optimum that a social planner would choose have largely focused on the extensive margin, addressing the question of whether the market leads to an under- or over-supply of varieties relative to the social optimum when preferences are additively separable. Dixit and Stiglitz (1977) provided the definitive answer to this question when firms are homogeneous: the market is efficient, in the sense that it supplies the socially optimal number of varieties, and the optimal output of each, if and only if preferences are CES. Feenstra and Kee (2008) showed that the market is also efficient with CES preferences if firms are heterogeneous and the distribution of firm productivities is Pareto, while Dhingra and Morrow (2019) present a general qualitative analysis of the heterogeneous-firm case. Here we focus on a quantitative comparison between the market outcome and the optimal allocation at the intensive margin. In particular, we show in Section 4.1 how our methods from previous sections can be used to derive closed-form expressions for the distributions of output in the competitive market equilibrium and in the social optimum. In Section 4.2 we compare the two distributions explicitly in the CREMR case, showing that, if and only if demand is subconvex, competitive markets encourage too many small firms and not enough large ones relative to the optimum. Other authors have derived related results in different contexts: e.g., Nocco et al. (2014), Edmond et al. (2015), and Behrens et al. (2020). However, these take different approaches from ours; in particular, they allow the extensive margin to adjust, and they do not compare the optimal and market output distributions directly as we do.

4.1 Equilibrium and Optimal Output Distributions

We wish to compare the market outcome with the social optimum. Consider first the former. Recalling from (3) that the first-order condition for each firm is that marginal cost should equal marginal revenue, so the productivity-output relationship is:

$$\varphi(x) = \frac{1}{r'(x)} = \frac{1}{p(x) + xp'(x)} \tag{24}$$

Letting $G(\varphi)$ denote the productivity distribution as before, this implies a new family of distributions, that we can call the "inverse marginal-revenue reflection" family:

$$J(x) = G(\varphi(x)) = G\left(\frac{1}{p(x) + xp'(x)}\right) \tag{25}$$

Using (25), we can compute the distribution of firm output in the market equilibrium $J(x)$ for any distribution of firm productivities $G(\varphi)$ and any demand function $p(x)$.

Consider next the social optimum. Following Dixit and Stiglitz (1977), we assume that the social planner cannot use lump-sum taxes or subsidies to affect profits. Extending the logic of this assumption to a heterogeneous-firms context, the feasible optimum is a constrained one, where the planner faces the same constraints as the market. In particular, she takes as given the mass of entrants, N_e, and the productivity threshold, $\underline{\varphi}$, equal to the productivity of a cutoff firm that makes zero profits in the market equilibrium. Given these, she maximizes aggregate utility:

$$\int_{i \in X} u(x(i))di = N_e \int_{\underline{\varphi}}^{\infty} u(x(\varphi))g(\varphi)d\varphi \tag{26}$$

where X is the set of goods produced, subject to the aggregate labor endowment constraint:[17]

$$N_e \int_{\underline{\varphi}}^{\infty} L\varphi^{-1} x(\varphi) g(\varphi) d\varphi + N_e f_e \leq L \tag{27}$$

The first-order condition for a social optimum is:

$$u'(x(\varphi)) = \lambda^* \varphi^{-1} \tag{28}$$

where λ^* is the shadow price of the constraint (27), which we can interpret as the social marginal utility of income; it is defined implicitly by (27) with equality and with $x(\varphi) = (u')^{-1}(\lambda^* \varphi^{-1})$. Hence the planner allocates production across firms according to:

$$\frac{u'(x(\varphi_i))}{u'(x(\varphi_j))} = \frac{\varphi_j}{\varphi_i} \tag{29}$$

which is a standard marginal-cost-pricing rule.

We can say more if the marginal utility of a threshold firm is finite: $u'(\underline{x}) < \infty$, where $\underline{x}$ is the output of a firm with productivity $\underline{\varphi}$. This could be because firms incur fixed costs, or because the demand function implies a finite upper bound for marginal revenue, as in the case of linear or strictly subconvex CREMR demands. Reexpressing (29) in terms of the output of a typical firm relative to that of a threshold one gives:

$$\frac{u'(x(\varphi))}{u'(\underline{x})} = \frac{\underline{\varphi}}{\varphi} \quad \Rightarrow \quad \varphi^*(x) = \underline{\varphi}\frac{u'(\underline{x})}{u'(x)} = \underline{\varphi}\frac{p(\underline{x})}{p(x)} \tag{30}$$

So the optimal productivity-output relationship depends only on demand (with $p(x)$ measuring the marginal willingness to pay at the optimum). This implies another new family of

[17] The number of firms that actually produce, and so the number of varieties available to consumers, is: $N = N_e \int_{\underline{\varphi}}^{\infty} g(\varphi) d\varphi = N_e(1 - G(\underline{\varphi}))$.

distributions that we call the "inverse marginal-utility reflection" family:

$$J^*(x) = G(\varphi^*(x)) = G\left(\underline{\varphi}\frac{u'(\underline{x})}{u'(x)}\right) = G\left(\underline{\varphi}\frac{p(\underline{x})}{p(x)}\right) \tag{31}$$

Just as we did for the market equilibrium, we can now compute the optimal distribution of output $J^*(x)$ for any distribution of firm productivities $G(\varphi)$ and any demand function $p(x)$.

4.2 Misallocation with CREMR Demands

To illustrate these general results, consider the distributions implied by CREMR demands. First we need the relationships between productivity and output in the market and socially optimal cases. These follow by using the expressions for CREMR marginal revenue and price in (24) and (30) respectively:

$$\varphi(x) = \frac{\sigma}{\beta(\sigma-1)}(x-\gamma)^{\frac{1}{\sigma}} = \underline{\varphi}\left(\frac{x-\gamma}{\underline{x}-\gamma}\right)^{\frac{1}{\sigma}} \quad \text{and} \quad \varphi^*(x) = \underline{\varphi}\,\frac{(\underline{x}-\gamma)^{\frac{\sigma-1}{\sigma}}}{\underline{x}}\frac{x}{(x-\gamma)^{\frac{\sigma-1}{\sigma}}} \tag{32}$$

The lower bounds for productivity and output are related in the same way as $\varphi(x)$ and x:

$$\underline{\varphi} = \varphi(\underline{x}) = \frac{\sigma}{\beta(\sigma-1)}(\underline{x}-\gamma)^{\frac{1}{\sigma}} \tag{33}$$

From (32), there is a simple relationship between the levels of productivity in the social optimum and the market equilibrium:

$$\varphi^*(x) = \frac{\underline{x}-\gamma}{\underline{x}}\frac{x}{x-\gamma}\varphi(x) \tag{34}$$

The coefficient of $\varphi(x)$ on the right-hand side is less than one if and only if γ is positive. Recalling that $J(x) = G(\varphi(x))$ and $J^*(x) = G(\varphi^*(x))$ yields a simple but important result:

Proposition 4. *Assume the distribution of firm productivity $G(\varphi)$ is continuous. Then, when demands are subconvex CREMR, the distribution of output in the social optimum $J^*(x)$*

first-order stochastically dominates that in the market equilibrium $J(x)$.

Heuristically, we can say that, with subconvex CREMR demands, the market equilibrium has too high a ratio of small to large firms relative to the social optimum.

It is straightforward to combine the productivity-output relationships from (32) with an assumed underlying productivity distribution in order to derive the distributions of output in the market equilibrium and the social optimum. We will see in Section 6.3 how these allow us to quantify the pattern of misallocation across firms, and to compare the social optimum and the market outcome at all points in the output distribution.

5 Inferring Sales and Markup Distributions

Next, we want to derive the distributions of sales r and markups $m \equiv \frac{p}{c}$, given the distribution of productivity and the form of the demand function. Section 5.1 shows how this is done in general; Section 5.2 considers the distributions of markups implied by CREMR demands; while Section 5.3 presents the distributions of both sales and markups implied by a number of widely-used demand functions.

5.1 Sales and Markup Distributions in General

In order to be able to invoke part (i) of Proposition 1, we need to express productivity as a function of sales and markups; combining these with the distribution of productivity allows us to derive the implied distributions of sales and markups: $F(r) = G(\varphi(r))$ and $B(m) = G(\varphi(m))$. To see how this works in practice, recall the relationship between productivity and output, $\varphi(x)$, from (24). (We illustrate for the case where the functional form of the inverse demand function, $p(x)$, is known. A similar approach is used when we know the direct demand function $x(p)$: see the discussion of the translog case in Appendix A.5.) Next, we need to relate output to sales and markups. For the former, we need to invert the function $r(x) = xp(x)$. When this can be done we can solve for $x(r)$, which gives $\varphi(r)$ by

substitution: $\varphi(r) = \varphi(x(r))$. For the latter, to express output as a function of the markup, we need to invert the function $m(x) = \frac{p(x)}{r'(x)}$. When this can be done, we again obtain $\varphi(m)$ by substitution: $\varphi(m) = \varphi(x(m))$.

5.2 CREMR Markup Distributions

To illustrate this approach, we consider the markup distributions implied by CREMR demands, which have the attraction that they take relatively simple forms. First, we can write the CREMR markup as a function of output: $m(x) = \frac{p(x)}{r'(x)} = \frac{x-\gamma}{x}\frac{\sigma}{\sigma-1}$. We concentrate on the case of strictly subconvex demands (i.e., $\gamma > 0$), which implies that larger firms have higher markups. Hence the support of the markup distribution is: $m(x) \in [\underline{m}, \overline{m})$; the minimum markup is $\underline{m} \equiv \frac{\underline{x}-\gamma}{\underline{x}}\frac{\sigma}{\sigma-1}$, where $\underline{x}$ is the minimum value of output, given by (33); while the upper bound of the markup, $\overline{m} \equiv \frac{\sigma}{\sigma-1}$, is the value that obtains under CES preferences with the same value of σ. Define the relative markup $\check{m}$ as the markup relative to its maximum value: $\check{m} \equiv \frac{m}{\overline{m}} = \frac{\sigma-1}{\sigma}m \in [\underline{\check{m}}, 1)$. Hence it follows that: $\check{m}(x) = \frac{x-\gamma}{x}$. Inverting this allows us to express output as a function of the relative markup: $x(\check{m}) = \frac{\gamma}{1-\check{m}}$. Finally, combining this with the CREMR relationship between productivity and output from (32), $\varphi(x)$, gives the desired relationship between productivity and the markup:

$$\varphi(\check{m}) = \frac{\underline{\varphi}}{\omega}\left(\frac{\check{m}}{1-\check{m}}\right)^{\frac{1}{\sigma}} \quad \text{where} \quad \omega \equiv \frac{\underline{\varphi}\beta}{\gamma^{\frac{1}{\sigma}}}\frac{\sigma-1}{\sigma} \tag{35}$$

From the discussion following Proposition 1, this implies that the distribution of markups is an "odds reflection" of that of productivity. Hence, if productivity follows any distribution in the GPF class and the demand function is subconvex CREMR, then relative markups follow the corresponding "GPF-odds" distribution. We illustrate for the Pareto and lognormal cases; extensions to other members of the GPF class are straightforward.[18]

[18]For example, if productivity follows a Fréchet distribution and demands are CREMR, the relative markup follows a "Fréchet-Odds" distribution, which provides an exact characterization of the distribution of profit margins for a firm selling in many foreign markets, when the distribution of productivity draws across markets follows a Fréchet distribution, as in Tintelnot (2017).

First, if productivity is distributed as a Pareto as in Corollary 1, then when demands are subconvex CREMR the relative markup has a "Pareto-Odds" distribution:

$$B(\check{m}) = G_{\mathcal{P}}(\varphi(\check{m})) = 1 - \left(\frac{\underline{\check{m}}}{1-\underline{\check{m}}}\right)^{n'} \left(\frac{\check{m}}{1-\check{m}}\right)^{-n'} \qquad \check{m} \in \{\underline{\check{m}}, 1\} \quad \check{m} \equiv \frac{m}{\overline{m}},\ \underline{\check{m}} \equiv \frac{\underline{m}}{\overline{m}}, \tag{36}$$

where $n' \equiv \frac{k}{\sigma}$ and $\underline{\check{m}} \equiv \frac{\omega^{\sigma}}{1-\omega^{\sigma}}$. This distribution appears to be new, and may prove useful in future applications.

Next, if productivity has a truncated lognormal distribution as in Corollary 2 and demands are subconvex CREMR, the relative markup has a "Lognormal-Odds" distribution:

$$B(\check{m}) = G_{t\mathcal{LN}}(\varphi(\check{m})) = \frac{\Phi\left(\frac{1}{\tilde{s}}\left(\log\frac{\check{m}}{1-\check{m}} - \tilde{\mu}\right)\right) - T}{1-T} \tag{37}$$

where: $\tilde{s} = \sigma s$, $\tilde{\mu} = \sigma\left(\mu + \log\left(\frac{\omega}{\underline{\varphi}}\right)\right)$, and $T = \Phi\left(\frac{1}{\tilde{s}}\left(\log\frac{\underline{\check{m}}}{1-\underline{\check{m}}} - \tilde{\mu}\right)\right)$ is the fraction of potential firms that are inactive, as in Corollary 2. This distribution has been studied in its untruncated form by Johnson (1949) and Mead (1965) who call it the "Logit-Normal", though we are not aware of a theoretical rationale for its occurrence as here. For some parameter values, it implies inverted-U-shaped markup densities similar to those found empirically by De Loecker et al. (2016) and Lamorgese et al. (2014), as discussed in the Introduction. In the next section we will compare these more formally.

5.3 Other Sales and Markup Distributions

Proposition 1 can be used to derive the distributions of sales and markups implied by any demand function. In particular, closed-form expressions for productivity as a function of sales or markups can be derived for some of the most widely-used demand functions in applied economics. Table 1 gives results for linear, LES, and translog demands, along with the CREMR results already derived in (4) and (35). Combining these with different assumptions about the distribution of productivity, and invoking Proposition 1, generates a wide variety of sales and markup distributions. For example, the relationships between productivity

	$p(x)$ or $x(p)$	$\varphi(r)$ or $\varphi(\check{r})$	$\varphi(m)$ or $\varphi(\check{m})$
CREMR	$\frac{\beta}{x}(x-\gamma)^{\frac{\sigma-1}{\sigma}}$	$\beta^{-\frac{\sigma}{\sigma-1}}\frac{\sigma}{\sigma-1}r^{\frac{1}{\sigma-1}}$	$\frac{1}{\beta}\frac{\sigma}{\sigma-1}\gamma^{\frac{1}{\sigma}}\left(\frac{\check{m}}{1-\check{m}}\right)^{\frac{1}{\sigma}}$
Linear	$\alpha-\beta x$	$\frac{1}{\alpha}\left(\frac{1}{1-\check{r}}\right)^{\frac{1}{2}}$	$\frac{2m-1}{\alpha}$
LES	$\frac{\delta}{x+\gamma}$	$\gamma\delta\left(\frac{1}{1-\check{r}}\right)^{2}$	$\frac{\gamma}{\delta}m^2$
Translog	$\frac{1}{p}(\gamma-\eta\log p)$	$(r+\eta)\exp\left(\frac{r-\gamma}{\eta}\right)$	$m\exp\left(m-\frac{\eta+\gamma}{\eta}\right)$

Table 1: Productivity as a Function of Sales and Markups for Selected Demand Functions

Note: See Appendix A.5 for detailed derivations. $\check{r}$ and $\check{m}$ denote sales and markups relative to their maximum values, respectively.

and sales implied by linear and LES demands have the same form, so the sales distributions implied by these two very different demand systems are observationally equivalent. The same is not true of their implied markup distributions, however; in the LES case, productivity is a simple power function of markups, so the LES implies self-reflection of the productivity and markup distributions if either is a member of the GPF class.[19] In the next section we compare the markup distributions implied by these different demand functions with each other and with a given empirical distribution.

6 Fitting Markups and Quantifying Misallocation

So far we have shown how to characterize the exact distributions of various firm outcomes (in particular firm sales, markups, output in the market equilibrium and socially-optimal output) implied by particular assumptions about the primitives of the model: the structure of demand and the distribution of firm productivities. In this section, we illustrate how, when applied to an actual data set, these theoretical results can be exploited empirically to estimate markup distributions and to quantify misallocation. In Section 6.1 we introduce the

[19]For example, a lognormal distribution of productivity and LES demands imply a lognormal distribution of markups, so providing microfoundations for an assumption made by Epifani and Gancia (2011).

firm-level data on Indian markups used in our econometric analysis. In Section 6.2, we fit the markup distributions implied by the theoretical models derived in Section 5 to the markup distributions in the data, and select the best-fitting models. For these models, we show in Section 6.3 how the approach to quantifying misallocation introduced in Section 4 can be implemented. In particular, we use the estimated parameter values obtained in Section 6.2 to infer the distributions of output given by the market and the one that would be chosen by the planner, and to compare them quantitatively.

As discussed in the introduction, Pareto and lognormal distributions yield very good fits for sales distributions. Thus, since CREMR demands exhibit self-reflection by construction, we would expect that, when combined with an underlying Pareto or lognormal distribution of productivities, they will yield a good fit to the distribution of sales.[20] By contrast, we are not aware of any previous attempts to fit the distribution of markups and explore its implications for misallocation in a theory-consistent way. Hence we focus in this section on fitting the distribution of markups.

6.1 The Data

The data set comes from De Loecker et al. (2016): see Appendix A.6 for more details. It consists of 2,457 firm-product observations on markups in Indian manufacturing for the year 2001. These markup data are estimated using the so-called "production approach": markups are calculated by computing the gap between the output elasticity with respect to variable inputs and the share of those inputs in total revenue.[21] This approach assumes cost minimization, a translog form for the production technology, and that some factor inputs are variable while others are fixed. However, it does not impose any restrictions on consumer demand nor on market structure. Hence it is particularly well-suited to our purpose, which is to compare the performance of different assumptions about the productivity distribution

[20] As we show in Online Appendices B.3 and B.6, this expectation is confirmed with data on Indian sales and French exports respectively, in line with previous literature.

[21] We do not have access to the confidence intervals for the markups and so we cannot take into account the fact that they were estimated, though it would be straightforward to do so.

and the demand function. The approach of De Loecker et al. (2016) has been criticised by Bond et al. (2020); however, the markup estimates we use are not subject to their main critique because they are based on output data rather than revenue data.

6.2 Actual Versus Predicted Markup Distributions

The approach we adopt builds directly on the theoretical framework developed in previous sections. Let $\tilde{B}(m)$ denote the markup distribution in the data, while $B(m;\boldsymbol{\theta})$ is the theory-consistent predicted distribution. $B(m;\boldsymbol{\theta})$ in turn is implied by an assumed underlying distribution of firm productivities, $G(\varphi,\boldsymbol{\theta}_1)$, combined with a productivity-markup relationship implied by an assumed demand function, $\varphi(m;\boldsymbol{\theta}_2)$, as given in Table 1:

$$B(m;\boldsymbol{\theta}) = G(\varphi(m,\boldsymbol{\theta}_2),\boldsymbol{\theta}_1), \tag{38}$$

where the parameter vector $\boldsymbol{\theta}$ is a function of the parameter vectors that characterize the productivity distribution and the demand function, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively. For each specification of G and φ, we estimate $\boldsymbol{\theta}$ that provides the best fit to the observed distribution $\tilde{B}(m)$. Note that in all cases $\boldsymbol{\theta}$ is of lower dimension than the combined dimensions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Hence, these parameters are not separately identified, so we cannot fully disentangle the effects of demand- and supply-side influences, though as we shall see we are able to discriminate between different demand functions given a maintained hypothesis about the productivity distribution.

To illustrate our approach in the simplest way, we select from the universe of potential specifications of productivity distributions and demand functions, a number of the most-widely-used alternatives which yield closed-form expressions for the implied distributions of firm markups and output. For the distribution of productivity, we focus on the Pareto and lognormal: both are plausible in themselves, albeit at different tails of the distribution, and they span a wide range of distributions that have been used in practice. As for our choice

of demand functions, we confine attention to the four demand functions presented in Table 1 in Section 5.3, all of which allow for variable markups.[22] It goes without saying that these choices represent only a limited selection from all possible specifications, but nonetheless a representative sample of current practice, especially when the constraints of tractability are taken into account.

Model	Markup PDF $b(m)$	Estimated Parameters	AIC	Relative Likelihood
CREMR $+\mathcal{P}$	$\frac{k\left(\frac{(\sigma-1)\underline{m}}{\underline{m}+\sigma-\underline{m}\sigma}\right)^{\frac{k}{\sigma}}}{(\sigma-1)m^2}\left(\frac{(\sigma-1)m}{m+\sigma-m\sigma}\right)^{\frac{\sigma-k}{\sigma}}$	$\sigma = 1.111$ $k = 1.233$	6049.43	
CREMR $+$ t$\mathcal{LN}$	$\frac{\frac{e^{-\frac{\left(\log\left(\frac{\sigma}{m+\sigma-m\sigma}-1\right)-\tilde{\mu}\right)^2}{2(\sigma s)^2}}}{\sqrt{2\pi}ms(m+\sigma-m\sigma)}}{1-\Phi\left(\frac{\log\left(\frac{(\sigma-1)\underline{m}}{\underline{m}+\sigma-\underline{m}\sigma}\right)-\tilde{\mu}}{\sigma s}\right)}$	$\tilde{\mu} = -49.982$ $s = 6.051$ $\sigma = 1.110$	6060.99	0.003
Linear $+t\mathcal{LN}$	$\sqrt{\frac{2}{\pi}}\frac{\frac{e^{-\frac{(\log(2m-1)-\tilde{\mu})^2}{2s^2}}}{s(2m-1)}}{1-\Phi\left(\frac{\log(2\underline{m}-1)-\tilde{\mu}}{s}\right)}$	$\tilde{\mu} = 0.054$ $s = 1.228$	6180.66	3.2×10^{-29}
LES $+t\mathcal{LN}$	$\sqrt{\frac{2}{\pi}}\frac{\frac{e^{-\frac{(2\log(m)-\tilde{\mu})^2}{2s^2}}}{ms}}{1-\Phi\left(\frac{2\log(\underline{m})-\tilde{\mu}}{s}\right)}$	$\tilde{\mu} = -3.234$ $s = 2.732$	6184.79	4.1×10^{-30}
LES $+\mathcal{P}$	$2k(\underline{m})^{2k}m^{-2k-1}$	$k = 0.747$	6244.63	4.1×10^{-43}
Translog $+\mathcal{P}$	$k(\underline{m}e^{\underline{m}})^k(m+1)m^{-k-1}e^{-km}$	$k = 0.487$	6258.08	4.9×10^{-46}
Translog $+t\mathcal{LN}$	$\frac{1}{\sqrt{2\pi}s}\frac{\frac{m+1}{m}e^{-\frac{\left(\log(me^m)-\tilde{\mu}\right)^2}{2s^2}}}{1-\Phi\left(\frac{\log(\underline{m}e^{\underline{m}})-\tilde{\mu}}{s}\right)}$	$\tilde{\mu} = -77.641$ $s = 13.064$	6283.10	1.8×10^{-51}
Linear $+\mathcal{P}$	$2k(2\underline{m}-1)^k\,(2m-1)^{-k-1}$	$k = 1.001$	6428.43	5.1×10^{-83}

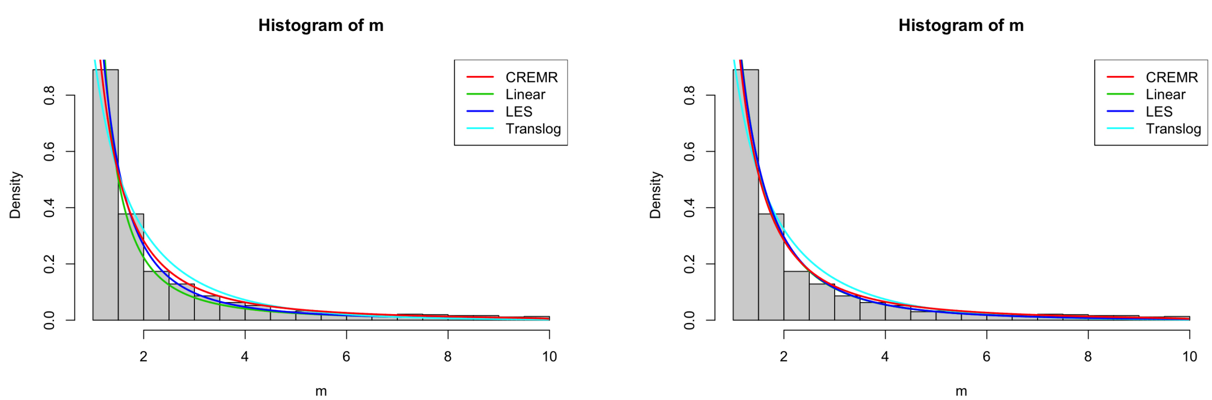
Table 2: Estimated Markup Densities Given Assumptions about Productivity (Pareto ($\mathcal{P}$) or Truncated Lognormal (t$\mathcal{LN}$)) and Demand (CREMR, Linear, LES or Translog).

Note: The ML estimator of the minimum markup $\underline{m}$ is the minimum empirical markup $m_{min} = 1.001$.

Using maximum-likelihood (ML) estimation, we fit the markup density functions implied

[22]We do not consider refinements of CES since they cannot match the heterogeneity of markups that we see in the data.

by eight different combinations of assumptions about the productivity distribution and the demand function. (Details of the estimation process and the code are available on our websites.) We have seen how to calculate the theoretical markup distributions in Section 5; further details are given in Appendix A.7. The results are summarized in Table 2 and illustrated in Figure 4. As the estimates in the third column show, some of the primitive parameters are identified: the Pareto shape parameter k, the lognormal standard deviation of the logs s, and the asymptotic CREMR demand elasticity σ. The other primitive parameters are not identified from our data, and are subsumed into the estimated parameters $\underline{m}$ and $\tilde{\mu}$: detailed expressions are given in Table 4 in Appendix A.7. To discriminate between different specifications, we use the Akaike information criterion (AIC), which is well-suited to compare models with different numbers of parameters, as it trades off goodness of fit and parameter parsimony. The models are ranked in the table by their AIC values as given in the second-last column. The final column gives the relative likelihood of the other models to the AIC-minimizing one, $\exp((AIC_{min} - AIC_i)/2)$, and can be interpreted as being proportional to the probability that the i'th model minimizes the estimated information loss.



(a) Pareto (b) Truncated Lognormal

Figure 4: Histograms of Empirical Markup Densities Compared with Fitted Densities Implied by Different Assumptions about Productivity (Pareto or Truncated Lognormal) and Demand (CREMR, Linear, LES or Translog)

Table 2 shows that the combination of Pareto productivity and CREMR demands minimizes the AIC. The truncated lognormal with CREMR model comes closest, but all the

others are far inferior by the AIC criterion. The Pareto assumption also gives a better fit with translog demands, but the truncated lognormal does better with LES and linear demands. These results suggest that the choice of productivity distribution is less important than the choice of demand function in fitting the data. Hence in the next subsection, we will explore misallocation focusing on the two best-fitting CREMR models.

6.3 Quantifying Misallocation

Next we want to use the estimated markup-distribution parameters from Section 6.2 for a quantitative comparison of the market and socially optimal distributions of output characterized in Section 4.2. As we have seen, not all the demand parameters are identified. Nonetheless, we can use the parameter estimates from the fitted markup distributions to illustrate the divergence between market and optimum, and we can exploit the properties of CREMR demands to draw some general conclusions.

Consider first the case where the productivity distribution is Pareto. As we saw in Section 4.2, we can combine this with the CREMR productivity-output relationship, using $J(x) = G(\varphi(x))$, to derive the implied distribution of output in the market equilibrium:

$$J(x) = 1 - \left(\frac{x-\gamma}{\underline{x}-\gamma}\right)^{-\frac{k}{\sigma}} = 1 - \gamma^{\frac{k}{\sigma}} \omega^k \left(x-\gamma\right)^{-\frac{k}{\sigma}} \tag{39}$$

The first expression depends on primitive parameters, of which two (k and σ) are observable using our data, while two ($\underline{x}$ and γ) are not; whereas in the second only γ is unobservable, since ω is a composite parameter that can be calculated from the estimates in Table 2. (See (36) and Appendix A.7.) The same holds for the optimal distribution of output with Pareto productivity:

$$J^*(x) = 1 - \left(\frac{(\underline{x}-\gamma)^{\frac{\sigma-1}{\sigma}}}{\underline{x}} \frac{x}{(x-\gamma)^{\frac{\sigma-1}{\sigma}}}\right)^{-k} = 1 - \gamma^{\frac{k}{\sigma}} \left(\frac{\omega^{\sigma-1}}{1+\omega^{\sigma}} x (x-\gamma)^{\frac{1-\sigma}{\sigma}}\right)^{-k} \tag{40}$$

If instead the productivity distribution is a lognormal, left-truncated at $\underline{\varphi}$, then the distribution of output in the market equilibrium is:

$$J(x) = \frac{\Phi\left(\frac{\log\left(\underline{\varphi}\left(\frac{x-\gamma}{\underline{x}-\gamma}\right)^{\frac{1}{\sigma}}\right)-\mu}{s}\right) - T}{1-T} = \frac{\Phi\left(\frac{\log\left(\frac{x-\gamma}{\gamma}\right)-\tilde{\mu}}{\sigma s}\right) - T}{1-T} \tag{41}$$

while the optimal distribution of output is:

$$J^{*}(x) = \frac{\Phi\left(\frac{\log\left(\underline{\varphi}\left(\frac{(\underline{x}-\gamma)^{\frac{\sigma-1}{\sigma}}}{\underline{x}}\frac{x}{(x-\gamma)^{\frac{\sigma-1}{\sigma}}}\right)\right)-\mu}{s}\right) - T}{1-T} = \frac{\Phi\left(\frac{\log\left(\frac{x^{\sigma}(x-\gamma)^{1-\sigma}}{\gamma}\right)-\tilde{\mu}_{x*}}{\sigma s}\right) - T}{1-T} \tag{42}$$

where $\tilde{\mu}_{x*} = \tilde{\mu} - \sigma \log\left(\frac{\sigma-1}{\sigma}\underline{m}\right)$ and T is the fraction of potential firms that are inactive, as in (37). Once again, conditional on γ, the final expressions in (41) and (42) are functions of observables (in this case $\tilde{\mu}$, s, σ, and $\underline{m}$).

Since γ is the only unobservable parameter in equations (39) to (42), we can illustrate the implied densities of output given the estimates of the other parameters from Table 2, conditional on different values of γ. Figures 5 and 6 do this for the Pareto and lognormal cases respectively; the output profiles in the market equilibrium are shown in blue and those in the social optimum in red. The differences between the two figures are not so surprising: for each value of γ, both the optimal and the market output profiles have more smaller firms in the lognormal case than in the long-tailed Pareto case. More surprising are the similarities between them: the densities are much more sensitive to the value of γ than to the specification of the productivity distribution; the optimal and market output densities intersect only once; and the value of output at which they intersect is increasing in γ. We will now prove these results more formally.

We have already seen in Proposition 4 that, with CREMR demands, the distribution of output in the social optimum first-order stochastically dominates that in the market equilibrium. We can go further and show that, conditional on observables, the critical value

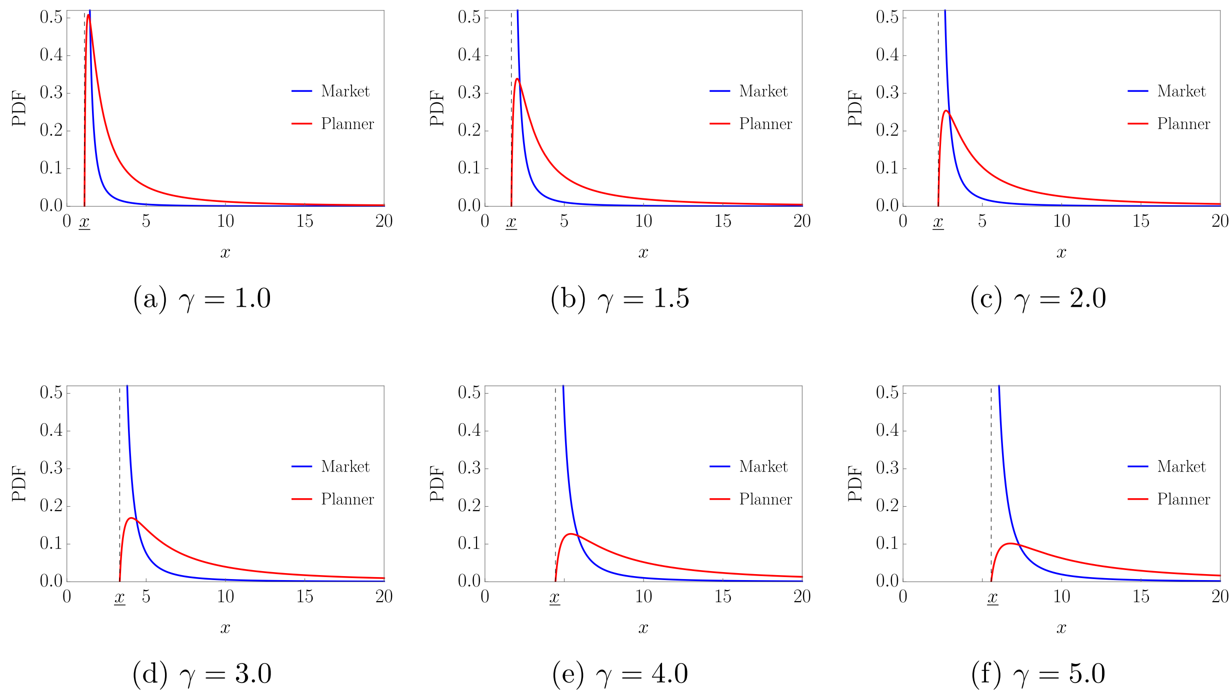


Figure 5: Market versus Socially-Optimal Output Profiles: Pareto and CREMR

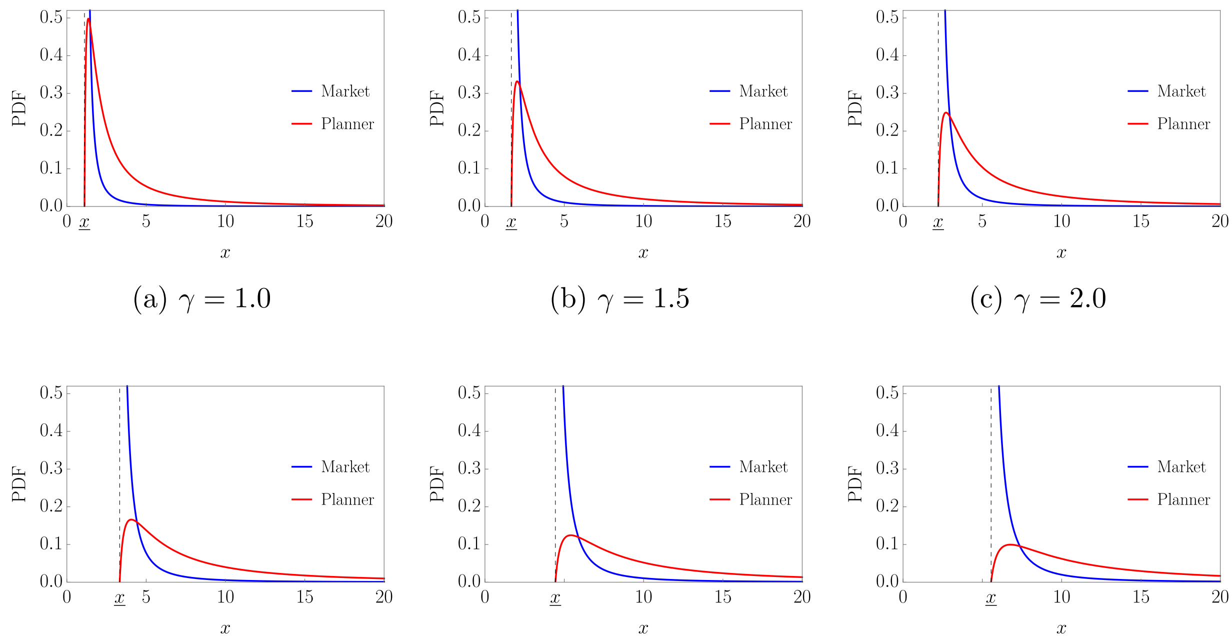
Note: Each curve shows the market (blue) or socially optimal (red) density of output, assuming a Pareto distribution of productivities and CREMR demands, with estimated parameters $k = 1.233$, $\sigma = 1.111$, and $\underline{m} = 1.001$, for different values of γ.

of output at which the market and optimal output densities intersect (which we denote by x_c) is unique and increases linearly in γ for both Pareto and lognormal distributions:

Lemma 1. *With subconvex CREMR demands, the critical value of output x_c at which the market and planner pdfs intersect is unique and proportional to γ, for any values of the observable parameters ($\underline{m}$, k, and σ) when the productivity distribution is Pareto, and for the estimated values of the observable parameters ($\underline{m}$, $\tilde{\mu}$, s and σ) when it is lognormal.*

The proof is in Appendix A.8. Using the estimated parameters from Table 2 we find that x_c equals 1.465γ in the Pareto case and 1.472γ in the lognormal case. Lemma 1 in turn implies that a key measure of misallocation does not depend on γ:

Proposition 5. *With CREMR demands and either Pareto or lognormal productivity, the fraction of firms that lies below the critical value of output x_c in both the market equilibrium and the social optimum is independent of γ for given observables: $\underline{m}$, k, and σ when the*



(a) $\gamma = 1.0$ (b) $\gamma = 1.5$ (c) $\gamma = 2.0$

(d) $\gamma = 3.0$ (e) $\gamma = 4.0$ (f) $\gamma = 5.0$

Figure 6: Market versus Socially-Optimal Output Profiles: Lognormal and CREMR

Note: Each curve shows the market (blue) or socially optimal (red) density of output, assuming a truncated lognormal distribution of productivities and CREMR demands, with estimated parameters $\tilde{\mu} = -49.982$, $s = 6.051$, $\sigma = 1.110$, and $\underline{m} = 1.001$, for different values of γ.

productivity distribution is Pareto, and $\underline{m}$, $\tilde{\mu}$, s and σ when it is lognormal.

The proof is in Appendix A.9. Using the estimated parameter values, we find that $J(x_c) = 0.795$ and $J^*(x_c) = 0.151$ in the Pareto case, while $J(x_c) = 0.797$ and $J^*(x_c) = 0.153$ in the truncated lognormal case. Hence we can conclude that, independent of γ, in the fitted CREMR-Pareto model, the market gives rise to $J(x_c)/J^*(x_c) = 5.252$ times as many "small" firms as a planner would choose, while the corresponding figure in the fitted CREMR-lognormal model is very similar: $J(x_c)/J^*(x_c) = 5.220$. Thus, although the key parameter γ is unobservable, we can draw a strong conclusion about the extent of misallocation with CREMR demands: for the parameters that we have estimated, the market equilibrium has over 5.2 times as many firms that are "too small" relative to the optimum.

7 Conclusion

This paper has addressed the question of how to relate the distributions of agent characteristics in models of heterogeneous agents. We provide a general necessary and sufficient condition for consistency between arbitrary assumptions about the distributions of two agent characteristics and an arbitrary behavioral model that relates those two characteristics at the individual level. In the specific context of Melitz-type models of heterogeneous firms competing in monopolistic competition, we showed that our condition implies a new demand function that generalizes the CES. The CREMR or "Constant Revenue Elasticity of Marginal Revenue" demand function is necessary and sufficient for the distributions of firm productivities and firm sales to be members of the same family. It is also necessary and sufficient for Gibrat's Law to hold over time under additive separability, it allows for variable markups in a parsimonious way, it provides a better empirical fit to data on Indian markups drawn from De Loecker et al. (2016) than a number of other better-known demand functions, and it leads to an operational way of quantifying the deviation of the competitive equilibrium from the social optimum. All these results hold for both Pareto and lognormal distributions of productivity, suggesting that the choice between these two distributions is less important than the choice between CREMR and other demands.

While we have concentrated on explaining the distributions of firm markups given assumptions about the distribution of firm productivity, it is clear that our approach has many other potential applications. As noted in the introduction, linking observed heterogeneity of outcomes to underlying heterogeneity of agents' characteristics via an assumed model of agent behavior is a common research strategy in many fields of economics. Both our general results and the specific functional forms we have introduced should prove useful in many other contexts.[23] To give just one example, the interaction of distributional and demand

[23] Other examples where our approach may prove fruitful are the interpretation of the trade elasticity, the elasticity of trade with respect to trade costs, which is a constant for many demand functions and Pareto productivities (see Arkolakis et al. (2012), Melitz and Redding (2015), and Arkolakis et al. (2018)) but not when the distribution of firm productivities is lognormal (see Head et al. (2014) and Bas et al. (2017)); and the granular origins of aggregate fluctuations, where Gabaix (2011) and di Giovanni and Levchenko (2012)

assumptions matters for quantifying the misallocation of resources, as we have shown in Sections 4 and 6. The pioneering study of Hsieh and Klenow (2009) estimated that close to half the difference in efficiency between China and India on the one hand and the U.S. on the other could be attributed to an inefficient allocation of labor and capital. However, this was under the maintained hypothesis that the output of each industry was a CES aggregate. As Dixit and Stiglitz (1977) and Feenstra and Kee (2008) showed, CES preferences for differentiated products imply that goods markets are constrained efficient. In a non-CES world, inefficiency may be partly a reflection of goods-market rather than factor-market distortions, with very different implications for welfare-enhancing policies.[24] In this and other cases, the assumptions made about the productivity distribution and demand structure matter for key economic issues, yet the existing literature gives little guidance on the implications of relaxing the standard assumptions, nor how best to proceed when the assumptions of the canonical model do not hold. Our paper has charted a way forward in both these directions.

have shown that relaxing the continuum assumption implies that the largest firms can have an impact on aggregate fluctuations, when the distribution of firm size is a power law in the upper tail.

[24]Arguments along these lines can be found, for example, in Epifani and Gancia (2011), Dhingra and Morrow (2019), and Haltiwanger et al. (2018).

A Appendix (For Publication)

This appendix gives technical details, proofs, and extensions. Sections A.1 and A.3 give more details on the class of Generalized Power Function distributions and the properties of CREMR demand functions respectively. Sections A.2, A.8, and A.9 give proofs of Proposition 1, Lemma 1 and Proposition 5 respectively. Section A.4 notes some further implications of Proposition 1, Section A.5 sketches the derivations underlying Table 1, Section A.6 discusses the data used in the empirics, while Section A.7 gives details of the markup distributions estimated in Section 6.2.

A.1 Generalized Power Function Distributions

Table 3 shows that many well-known distributions are members of the Generalized Power Function class, $G(y;\boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2} y^{\theta_2}\right)$, introduced in Definition 1. Hence Proposition 1 can immediately be applied to deduce a constant-elasticity relationship between any two firm characteristics which share any of the distributions in the table, provided the two distributions have compatible supports, and the same value of the parameter θ_0.

	$G(y;\boldsymbol{\theta})$	Support	$H(z)$	θ_0	θ_1	θ_2
Pareto	$1-\underline{y}^k y^{-k}$	$[\underline{y},\infty)$	z	1	$k\underline{y}^k$	$-k$
Right-Truncated Pareto	$\frac{1-\underline{y}^k y^{-k}}{1-\underline{y}^k \overline{y}^{-k}}$	$[\underline{y},\overline{y}]$	z	$\frac{1}{1-\underline{y}^k\overline{y}^{-k}}$	$\frac{k\underline{y}^k}{1-\underline{y}^k\overline{y}^{-k}}$	$-k$
Lognormal	$\Phi((\log y-\mu)/s)$	$[0,\infty)$	$\Phi(\log z)$	0	$\frac{1}{s}\exp\left(-\frac{\mu}{s}\right)$	$\frac{1}{s}$
Left-Truncated Lognormal*	$\frac{\Phi((\log y-\mu)/s)-T}{1-T}$	$[\underline{y},\infty)$	$\frac{\Phi(\log z)-T}{1-T}$	0	$\frac{1}{s}\exp\left(-\frac{\mu}{s}\right)$	$\frac{1}{s}$
Uniform	$\frac{y-\underline{y}}{\overline{y}-\underline{y}}$	$[\underline{y},\overline{y}]$	z	$-\frac{\underline{y}}{\overline{y}-\underline{y}}$	$\frac{1}{\overline{y}-\underline{y}}$	1
Fréchet	$\exp\left(-\left(\frac{y-\mu}{s}\right)^{-\alpha}\right)$	$[\mu,\infty)$	$\exp(-z^{-\alpha})$	$-\frac{\mu}{s}$	$\frac{1}{s}$	1
Gumbel	$\exp\left(-\exp\left(-\left(\frac{y-\mu}{s}\right)\right)\right)$	$(-\infty,\infty)$	$\exp(-\exp(-z))$	$-\frac{\mu}{s}$	$\frac{1}{s}$	1
Reversed Weibull	$\exp\left(-\left(\frac{\mu-y}{s}\right)^{\alpha}\right)$	$(-\infty,\mu]$	$\exp(-z^{\alpha})$	$\frac{\mu}{s}$	$-\frac{1}{s}$	1

Table 3: Some Members of the Generalized Power Function Class of Distributions

Note: $^{*}T = \Phi\left((\log \underline{y}-\mu)/s\right)$.

A useful result is that an arbitrarily truncated member of the GPF class is also a member:

Corollary 3. *If the family of distributions $G(y;\boldsymbol{\theta})$ is a member of the GPF class, then the family $\tilde{G}(y;\underline{y},\overline{y},\boldsymbol{\theta})$, formed by truncating $G(y;\boldsymbol{\theta})$ to the interval $y \in [\underline{y},\overline{y}]$, is also a member of the GPF class.*

The corollary follows immediately from the properties of a truncated distribution:

$$\tilde{G}(y;\underline{y},\overline{y},\boldsymbol{\theta}) = \tilde{H}\left(\theta_0 + \frac{\theta_1}{\theta_2} y^{\theta_2}\right) \qquad \text{where:} \qquad \tilde{H}(z) \equiv \frac{H(z) - G(\underline{y};\boldsymbol{\theta})}{G(\overline{y};\boldsymbol{\theta}) - G(\underline{y};\boldsymbol{\theta})} \tag{43}$$

When comparing two distributions from the same family, their supports must be compatible. For example, if the distributions of y and z are from the same truncated family that is a member of the GPF class, then we must have $G(\underline{y};\boldsymbol{\theta}) = G(\underline{z};\boldsymbol{\theta})$ and $G(\overline{y};\boldsymbol{\theta}) = G(\overline{z};\boldsymbol{\theta})$.

A simple example of a distribution that is not a member of the GPF class is the exponential: $G(y;\boldsymbol{\theta}) = 1 - \exp(-\lambda y)$. This one-parameter distribution does not have the flexibility to match either the sufficiency or the necessity part of Proposition 2. If y is distributed as an exponential and $y = y_0 z^E$, then z is distributed as a Weibull: $F(z;\boldsymbol{\theta'}) = 1 - \exp(-\lambda y_0 z^E)$. Whereas if both y and z are distributed as exponentials, then $y = y_0 z$, i.e., $E = 1$. For similar reasons, the one-parameter version of the Fréchet (used in Eaton and Kortum (2002)) is not a member of the GPF class, though as Table 3 shows, both its two-parameter version (used in many applications of the Eaton-Kortum model) and the three-parameter "Translated Fréchet" (with one of the parameters equal to θ_0) are members.

A.2 Proof of Proposition 1

We first give a preliminary lemma, that characterizes the links between the distributions of y and z and the theoretical function linking them:[25]

Lemma 2. *Given Assumption 1, any two of the following imply the third:*

(A) y is distributed with CDF $G(y)$, where $g(y) \equiv G'(y) > 0$;

[25] We consider a continuum of agents whose characteristics are realizations of a random variable. Because we work with a continuum, the c.d.f. of this random variable is the actual distribution of these realizations. Henceforward, we use lower-case variables to describe both a random variable and its realization.

(B) z is distributed with CDF $F(z)$, where $f(z) \equiv F'(z) > 0$;

(C) Agent behavior is such that: $y = v(z)$, $v'(z) > 0$;

where the functions are related as follows:

(i) (A) and (C) imply (B) with $F(z) = G(v(z))$ and $f(z) = g(v(z))v'(z)$; similarly, (B) and (C) imply (A) with $G(y) = F(v^{-1}(y))$ and $g(y) = f(v^{-1}(y))\frac{d(v^{-1}(y))}{dy}$.

(ii) (A) and (B) imply (C) with $v(z) = G^{-1}(F(z))$.

Part (i) of the lemma is a standard result on transformations of variables. Part (ii) is less standard (it is closely related to Lemma 1 of Matzkin (2003)), and requires Assumption 1: characteristics $y(i)$ and $z(i)$ must refer to the same agent and must be monotonically increasing in i. The importance of the result is that it allows us to characterize fully the conditions under which assumptions about the distributions of two variables and about the relationship that links them are mutually consistent. Part (ii) in particular provides an easy way of determining which specifications of agent behavior are consistent with particular assumptions about the distributions of agent characteristics. All that is required is to derive the form of $v(z)$ implied by any pair of distributional assumptions.

As already noted, part (i) of Lemma 2 is standard. To prove part (ii), that (A) and (B) imply (C), consider an arbitrary firm i with characteristics $y(i)$ and $z(i)$. Because $y(i)$ and $z(i)$ are monotonically increasing in i, the fraction of firms with characteristics equal to or less than $y(i)$ and $z(i)$, are equal:

$$G[y(i)] = F[z(i)] \quad \forall i \in \Omega \tag{44}$$

Inverting gives $y(i) = G^{-1}[F(z(i))]$. Since this holds for any firm $i \in \Omega$, it follows that $y = v(z) = G^{-1}[F(z)]$, as required. This completes the proof of Lemma 2.

Next, we turn to the proof of Proposition 1 itself. The proof of part (i) is immediate. To show that (A) and (C) imply (B), assume $G(y; \boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2} y^{\theta_2}\right)$, $G_y > 0$, and

$y = y_0 h(z)^E$. Then the implied distribution of z is:

$$F\left(z; \boldsymbol{\theta}\right) = H\left(\theta_0 + \frac{\theta_1}{\theta_2}\left(y_0 h(z)^E\right)^{\theta_2}\right) = H\left(\theta_0 + \frac{\theta_1'}{\theta_2'} h(z)^{\theta_2'}\right) \tag{45}$$

where: $\theta_2' = E\theta_2$ and $\frac{\theta_1'}{\theta_2'} = \frac{\theta_1}{\theta_2} y_0^{\theta_2}$ so $\theta_1' = \frac{\theta_1}{\theta_2}\theta_2' y_0^{\theta_2} = E\theta_1 y_0^{\theta_2}$. Thus (A) and (C) imply (B). A similar proof shows that (B) and (C) imply (A).

Next, we wish to prove part (ii), that (A) and (B) imply (C). The proof uses a key property of the GPF class: the composition of one member of a GPF family and the inverse of another is a power function: $G^{-1}\left(G(z, \boldsymbol{\theta'}), \boldsymbol{\theta}\right) = y_0 z^E$. Let ψ equal the distribution in (A): $\psi \equiv G\left(y; \boldsymbol{\theta}\right) = H\left(\theta_0 + \frac{\theta_1}{\theta_2} y^{\theta_2}\right)$, $G_y > 0$. Inverting the G and H functions gives: $y \equiv G^{-1}\left(\psi; \boldsymbol{\theta}\right)$ and $\theta_0 + \frac{\theta_1}{\theta_2} y^{\theta_2} = H^{-1}(\psi)$. Eliminating y yields: $G^{-1}\left(\psi; \boldsymbol{\theta}\right) = \left(\frac{\theta_2}{\theta_1}\left(H^{-1}(\psi) - \theta_0\right)\right)^{\frac{1}{\theta_2}}$. Invoking part (ii) of Lemma 2, we can state that $y = G^{-1}\left(F\left(z; \boldsymbol{\theta'}\right); \boldsymbol{\theta}\right)$, where $F\left(z; \boldsymbol{\theta'}\right) = H\left(\theta_0 + \frac{\theta_1'}{\theta_3'} h(z)^{\theta_2'}\right)$, $F_z > 0$, from (B). Hence, substituting for $G^{-1}\left(\psi; \boldsymbol{\theta}\right)$ and $F\left(z; \boldsymbol{\theta'}\right)$ gives:

$$y = \left(\frac{\theta_2}{\theta_1}\left(H^{-1}\left(H\left(\theta_0 + \frac{\theta_1'}{\theta_3'} h(z)^{\theta_2'}\right)\right) - \theta_0\right)\right)^{\frac{1}{\theta_2}} = \left(\frac{\theta_2}{\theta_1}\left(\left(\theta_0 + \frac{\theta_1'}{\theta_2'} h(z)^{\theta_2'}\right) - \theta_0\right)\right)^{\frac{1}{\theta_2}} = y_0 h(z)^E \tag{46}$$

where: $E = \frac{\theta_2'}{\theta_2}$ and $y_0 = \left(\frac{\theta_2}{\theta_1}\frac{\theta_1'}{\theta_3'}\right)^{\frac{1}{\theta_2}} = \left(\frac{1}{E}\frac{\theta_1'}{\theta_1}\right)^{\frac{1}{\theta_2}}$. Thus (A) and (B) imply (C), which completes the proof of Proposition 1.

A.3 Properties of CREMR Demand Functions

First, we wish to show that the CREMR property $\varphi = (r')^{-1} = \varphi_0 r^E$ is necessary and sufficient for the CREMR demands given in (6). To prove sufficiency, note that, from (6), total and marginal revenue given CREMR demands are:

$$r(x) \equiv xp(x) = \beta\left(x - \gamma\right)^{\frac{\sigma-1}{\sigma}} \qquad r'(x) = p(x) + xp'(x) = \beta\frac{\sigma - 1}{\sigma}\left(x - \gamma\right)^{-\frac{1}{\sigma}} \tag{47}$$

Combining these, the revenue elasticity of marginal revenue is constant, equal to $\frac{1}{\sigma-1}$:

$$r'(x) = \beta^{\frac{\sigma}{\sigma-1}} \, \frac{\sigma-1}{\sigma} \, r(x)^{-\frac{1}{\sigma-1}} \tag{48}$$

To prove necessity, invert equation (5) to obtain $r'(x) = \varphi_0^{-1} r(x)^{-E}$. This is a standard first-order differential equation in $r(x)$ with constant coefficients. Its solution is:

$$r(x) = \left((E+1)\left(\varphi_0^{-1} x - \kappa\right)\right)^{\frac{1}{E+1}} \tag{49}$$

where κ is a constant of integration. Collecting terms, recalling that $r(x) = xp(x)$, gives the CREMR demand system (6), where the coefficients are: $\sigma = \frac{E+1}{E}$, $\beta = (E+1)^{\frac{1}{E+1}} \varphi_0^{-\frac{1}{E+1}}$, and $\gamma = \varphi_0 \kappa$. Note that it is the constant κ which makes CREMR more general than CES. Since the CREMR property $\varphi = (r')^{-1} = \varphi_0 r^E$ is both necessary and sufficient for the demands given in (6), we call the latter CREMR demands.

Next, we wish to derive the demand manifold for CREMR demand functions. This can be done directly by calculating the elasticity and convexity of demand:

$$\varepsilon = \frac{x-\gamma}{x-\gamma\sigma}\sigma \qquad \text{and} \qquad \rho = 2 - \left(\frac{1}{x-\gamma\sigma} - \frac{1}{(x-\gamma)\sigma}\right) x \tag{50}$$

Eliminating x yields the CREMR demand manifold in the text, equation (21). An alternative route to deriving this result that proves useful in Section 3.3 follows Mrázová and Neary (2017), who show that, for a firm with constant marginal cost facing an arbitrary demand function, the elasticities of total and marginal revenue with respect to output can be expressed in terms of the elasticity and convexity of demand. These results yield an expression for the revenue elasticity of marginal revenue which holds for any demand function:

$$\left.\begin{aligned} \widehat{r} &= \tfrac{\varepsilon-1}{\varepsilon}\,\widehat{x} \\ \widehat{r'} &= -\tfrac{2-\rho}{\varepsilon-1}\,\widehat{x} \end{aligned}\right\} \quad \Rightarrow \quad \widehat{r'} = -\frac{\varepsilon(2-\rho)}{(\varepsilon-1)^2}\,\widehat{r} \tag{51}$$

Equating the coefficient of $\widehat{r}$ to the CREMR elasticity from (49) above, again leads to equation (21). Note that requiring marginal revenue to be positive ($\varepsilon > 1$) and decreasing ($\rho < 2$) implies that $\sigma > 1$, just as in the familiar CES case.

The expression for the elasticity of demand in (50) shows that it decreases in output, and so demand is subconvex, if and only if γ is positive. Further details on this and other properties of CREMR demands are given in Online Appendix B.2.

A.4 Other Implications of Proposition 1

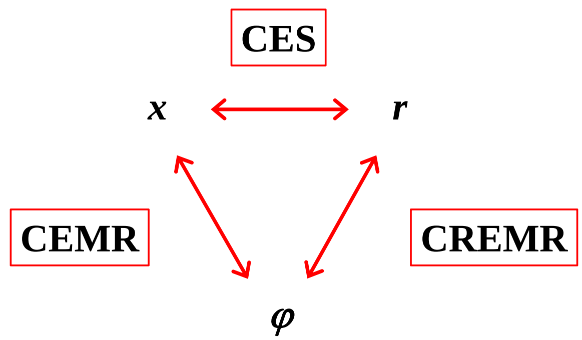


Figure 7: Demand Functions that Yield Self-Reflection
Between the Distributions of Firm Characteristics

Section 3 showed that CREMR demands are necessary and sufficient for the distributions of productivity φ and sales revenue r to be members of the same family of the GPF class of distributions. In Online Appendix B.1 we use the same approach to characterize the demand functions that are necessary and sufficient for self-reflection between the distributions of productivity and firm output x, and between output and sales revenue. Figure 7 summarizes schematically all these results: necessary and sufficient for self-reflection of the distributions of productivity and output is a demand function that we call “CEMR” (Constant Elasticity of output with respect to Marginal Revenue); while CES demands are necessary and sufficient for self-reflection of the distributions of output and sales. Corollaries of these results are that CES is sufficient for each of the three bilateral links between distributions, and is necessary and sufficient for all three distributions to have the same form.

A.5 Derivations Underlying Table 1

As in Mrázová and Neary (2017), we give the demand functions from a "firm's-eye view"; many of the parameters taken as given by the firm are endogenous in industry equilibrium. For each demand function, we follow a similar approach to that used with CREMR demands in Sections 3.2 and 5.2: we use the first-order condition to solve for productivity as a function of either output or price; the definition of sales revenue to solve for output or price as a function of sales; and the relationship between markups and elasticities to solve for output or price as a function of the markup. Combining yields $\varphi(r)$ and $\varphi(m)$ as required.

Linear: $p(x) = \alpha - \beta x$, $\alpha > 0, \beta > 0$. Sales revenue is quadratic in output, $r(x) = \alpha x - \beta x^2$, but only the root corresponding to positive marginal revenue, $r'(x) = \alpha - 2\beta x > 0$, is admissible. Since maximum output is $\overline{x} = \frac{\alpha}{2\beta}$, maximum sales revenue is $\overline{r} = \frac{\alpha^2}{4\beta}$, and we work with sales relative to their maximum: $\check{r} \equiv \frac{r}{\overline{r}}$. Hence output as a function of relative sales is: $x(\check{r}) = \frac{\alpha}{2\beta}\left(1 - (1-\check{r})^{\frac{1}{2}}\right)$. Equating marginal revenue to marginal cost gives $\varphi(x) = \frac{1}{\alpha - 2\beta x}$, so we can calculate $\varphi(\check{r}) = \varphi(x(\check{r}))$. As for the markup, as a function of output it is $m(x) = \frac{p(x)}{r'(x)} = \frac{\alpha - \beta x}{\alpha - 2\beta x}$. We do not work with the relative markup in this case, since $m(x) \to \infty$ as $x \to \overline{x}$. Inverting $m(x)$ gives $x(m) = \frac{\alpha}{\beta}\frac{m-1}{2m-1}$, from which we can calculate $\varphi(m) = \varphi(x(m))$ in Table 1.

LES: $p(x) = \frac{\delta}{x+\gamma}$, $\gamma > 0, \delta > 0$. We use the inverse demand function rather than the more familiar direct one: $x(p) = \frac{\delta}{p} - \gamma$. In monopolistic competition, the second-order condition requires that γ be positive, which rules out its usual interpretation as minus a subsistence level of consumption and also guarantees subconvexity. Sales revenue is $r(x) = \delta\frac{x}{x+\gamma}$, attaining its maximum at $\overline{r} = \delta$, so we work with relative sales: $\check{r} \equiv \frac{r}{\overline{r}} = \frac{x}{x+\gamma}$. Inverting gives: $x(\check{r}) = \gamma\frac{\check{r}}{1-\check{r}}$. The first-order condition yields: $\varphi(x) = \frac{(x+\gamma)^2}{\gamma\delta}$. Combining gives $\varphi(\check{r}) = \varphi(x(\check{r}))$. Finally, the markup as a function of output is $m(x) = \frac{p(x)}{r'(x)} = \frac{x+\gamma}{\gamma}$; inverting gives $x(m) = \gamma(m-1)$, which again yields $\varphi(m) = \varphi(x(m))$.

Translog: $x(p) = \frac{1}{p}(\gamma - \eta\log p)$, $\gamma > 0, \eta > 0$. From the direct demand function, sales revenue as a function of price is $r(p) = \gamma - \eta\log p$. Inverting gives $p(r) = \exp\left(\frac{\gamma - r}{\eta}\right)$. From the

first-order condition, $\varphi(p) = \frac{x'(p)}{r'(p)} = \frac{\eta+\gamma-\eta\log p}{\eta p}$. Combining this with $p(r)$ gives the expression for $\varphi(r)$ in Table 1. Finally, the markup as a function of price is $m(p) = \frac{p}{r'(p)} = \frac{\eta+\gamma-\eta\log p}{\eta}$; inverting gives $p(m) = \exp\left(\frac{\eta+\gamma}{\eta} - m\right)$, which yields $\varphi(m) = \varphi(p(m))$ in Table 1.

A.6 Data on Indian Sales and Markups

See De Loecker et al. (2016) for a detailed description of the data, which come from the Prowess data set collected by the Centre for Monitoring the Indian Economy (CMIE). Observations with negative markups (about 20% of the total) are not included in the sample, as they are inconsistent with steady-state equilibrium behavior by firms.

A.7 Markup Distributions

Demand Function	**Pareto Productivity** $G_{\mathcal{P}}(\varphi) = 1 - \underline{\varphi}^k \varphi^{-k}$	**Truncated Lognormal Productivity** $G_{t\mathcal{LN}}(\varphi) = \frac{\Phi\left(\frac{\log\varphi-\mu}{s}\right)-T}{1-T}$
CREMR $p(x) = \frac{\beta}{x}\left(x-\gamma\right)^{\frac{\sigma-1}{\sigma}}$	$B(m) = 1 - \omega^k \left(\frac{(\sigma-1)m}{m+\sigma-\sigma m}\right)^{-\frac{k}{\sigma}}$ $\omega = \frac{\varphi\beta}{\gamma^{\frac{1}{\sigma}}} \frac{\sigma-1}{\sigma} = \left(\frac{(\sigma-1)\underline{m}}{\underline{m}+\sigma-\sigma\underline{m}}\right)^{\frac{1}{\sigma}}$	$B(m) = \frac{\Phi\left(\frac{\log\left(\frac{(\sigma-1)m}{m+\sigma-m\sigma}\right)-\tilde{\mu}}{s\sigma}\right)-T}{1-T}$ $\tilde{\mu} = \sigma\left(\mu - \log\left(\frac{\sigma}{\sigma-1}\frac{\gamma^{\frac{1}{\sigma}}}{\beta}\right)\right)$ $T = \Phi\left(\frac{\log\underline{\varphi}-\mu}{s}\right) = \Phi\left(\frac{\log\left(\frac{(\sigma-1)\underline{m}}{\underline{m}+\sigma-\underline{m}\sigma}\right)-\tilde{\mu}}{s\sigma}\right)$
Linear $p(x) = \alpha - \beta x$	$B(m) = 1 - \left(\frac{2m-1}{\omega}\right)^{-k}$ $\omega = \alpha\underline{\varphi} = 2\underline{m} - 1$	$B(m) = \frac{\Phi\left(\frac{\log(2m-1)-\tilde{\mu}}{s}\right)-T}{1-T}$ $\tilde{\mu} = \mu + \log\alpha$ $T = \Phi\left(\frac{\log\underline{\varphi}-\mu}{s}\right) = \Phi\left(\frac{\log(2\underline{m}-1)-\tilde{\mu}}{s}\right)$
LES $p(x) = \frac{\delta}{x+\gamma}$	$B(m) = 1 - \left(\frac{m^2}{\omega}\right)^{-k}$ $\omega = \frac{\delta\underline{\varphi}}{\gamma} = \underline{m}^2$	$B(m) = \frac{\Phi\left(\frac{2\log(m)-\tilde{\mu}}{s}\right)-T}{1-T}$ $\tilde{\mu} = \mu + \log(\frac{\delta}{\gamma})$ $T = \Phi\left(\frac{\log\underline{\varphi}-\mu}{s}\right) = \Phi\left(\frac{2\log(\underline{m})-\tilde{\mu}}{s}\right)$
Translog $x(p) = \frac{1}{p}(\gamma - \eta\log p)$	$B(m) = 1 - \omega^k \left(me^m\right)^{-k}$ $\omega = e^{1+\frac{\gamma}{\eta}}\underline{\varphi} = \underline{m}e^{\underline{m}}$	$B(m) = \frac{\Phi\left(\frac{\log(me^m)-\tilde{\mu}}{s}\right)-T}{1-T}$ $\tilde{\mu} = \mu + 1 + \frac{\gamma}{\eta}$ $T = \Phi\left(\frac{\log\underline{\varphi}-\mu}{s}\right) = \Phi\left(\frac{\log(\underline{m}e^{\underline{m}})-\tilde{\mu}}{s}\right)$

Table 4: Markup Distributions Implied by Assumptions about Productivity (Pareto ($\mathcal{P}$) or Truncated Lognormal (t$\mathcal{LN}$)) and Demand (CREMR, Linear, LES or Translog).

Table 4 gives the markup distributions implied by different assumptions about the underlying productivity distribution and the demand function, using the relation $B(m) = G(\varphi(m))$. The two distributions implied by CREMR demands have already been given in Section 5.2 in terms of the relative markup as a function of primitive parameters. (Recall equations (36) and (37).) Here all eight distributions are expressed in terms of observables; the estimated values of these are given in Table 2 in the text. The expressions are simplified by writing them in terms of composite parameters ω and T which can be calculated from observable parameters. The table also shows how all observable parameters can be expressed in terms of unobservable primitive parameters.

A.8 Proof of Lemma 1

Begin with the case of a Pareto productivity distribution. We first show that the market equilibrium and optimal output pdfs cross only once. Using the distributions (39) and (40), we derive the densities $j(x)$ and $j^*(x)$. At any point where these intersect:

$$\gamma^{\frac{k}{\sigma}} \frac{k}{\sigma} \omega^k (x-\gamma)^{-\frac{k+\sigma}{\sigma}} = \gamma^{\frac{k}{\sigma}} \frac{k}{\sigma} \frac{x-\gamma\sigma}{x(x-\gamma)} \left(\frac{\omega^{\sigma-1}}{1+\omega^{\sigma}} x (x-\gamma)^{\frac{1-\sigma}{\sigma}} \right)^{-k} \tag{52}$$

$$\Leftrightarrow \quad \left(\frac{\omega^{\sigma}}{1+\omega^{\sigma}} \right)^k = \frac{x-\gamma\sigma}{x} \left(\frac{x-\gamma}{x} \right)^k \tag{53}$$

Given $x > \gamma\sigma > \gamma$, the right-hand side of (53) is increasing in x. Two densities must intersect at least once, so this proves that they intersect only once. Next, we hypothesize that, for given values of ω, k and σ, the unique solution is linear in γ, so $x_c = a\gamma$, where a is a constant. Substituting this into (53), γ cancels, leaving an implicit expression for a as a function of ω, k and σ only. Recalling that ω is a composite parameter which can be expressed in terms of $\underline{m}$ and σ as in (36), this confirms that, for given values of observables $\underline{m}$, k, and σ, $x_c = a\gamma$ is the unique solution of $j(x) = j^*(x)$, which proves Lemma 1 for the Pareto case.

The proof in the case of a lognormal productivity distribution proceeds in the same way.

Using the distributions (41) and (42), we derive the densities. At an intersection point:

$$\frac{\frac{\exp\left(-\frac{\left(\log\left(\frac{x-\gamma}{\gamma}\right)-\tilde{\mu}\right)^2}{2s^2\sigma^2}\right)}{\sqrt{2\pi}s\sigma(x-\gamma)}}{1-T} = \frac{\frac{(x-\gamma\sigma)\exp\left(-\frac{\left(\log\left(\frac{x^\sigma(x-\gamma)^{1-\sigma}}{\gamma}\right)-\tilde{\mu}x*\right)^2}{2s^2\sigma^2}\right)}{\sqrt{2\pi}s\sigma x(x-\gamma)}}{1-T} \tag{54}$$

$$\Leftrightarrow \quad 1 = \frac{x-\gamma\sigma}{x}\left(\frac{1+\omega^\sigma}{\omega^\sigma}\frac{x-\gamma}{x}\right)^{\frac{\log\left(\left(\frac{\omega^\sigma}{1+\omega^\sigma}\right)^\sigma \frac{x^\sigma(x-\gamma)^{2-\sigma}}{\gamma^2}\right)-2\tilde{\mu}}{2s^2\sigma}} \tag{55}$$

Recalling that $\omega = \left(\frac{(\sigma-1)\underline{m}}{\underline{m}+\sigma-\sigma\underline{m}}\right)^{\frac{1}{\sigma}}$, for the estimated values of the observable parameters $\tilde{\mu}$, σ, s and $\underline{m}$, and any γ, the right-hand side of (55) is monotonically increasing in x for $x > \gamma\sigma > \gamma$. This proves that the densities intersect only once. Next, we test the solution $x_c = a\gamma$ where a is a constant for the estimated parameters. Substituting into (55) yields an implicit expression for a as a function of $\tilde{\mu}$, σ, s and ω. This confirms that, for the estimated values of observable parameters $\tilde{\mu}$, σ, s and $\underline{m}$, $x_c = a\gamma$ is the unique solution of $j(x) = j^*(x)$, which proves Lemma 1 for the lognormal case.

A.9 Proof of Proposition 5

Consider first the Pareto case. Evaluating (39) and (40) at $x = x_c = a\gamma$ yields:

$$J(x_c) = 1 - \omega^k (a-1)^{-\frac{k}{\sigma}} \quad \text{and} \quad J^*(x_c) = 1 - \left(\frac{\omega^{\sigma-1}}{1+\omega^\sigma}a(a-1)^{\frac{1-\sigma}{\sigma}}\right)^{-k} \tag{56}$$

Recalling that ω is a composite parameter which can be expressed in terms of $\underline{m}$ and σ as in (36), this confirms that, for given values of observables $\underline{m}$, k, and σ, the proportion of firms that are smaller or equal to x_c is independent of γ in both the market and the planner's distributions.

Similarly, in the lognormal case, evaluating (41) and (42) at $x = x_c = a\gamma$ yields:

$$J(x_c) = \frac{\Phi\left(\frac{\log(a-1)-\tilde{\mu}}{\sigma s}\right) - T}{1-T} \quad \text{and} \quad J^*(x_c) = \frac{\Phi\left(\frac{\log\left((a-1)^{1-\sigma}\right)-\tilde{\mu}_{x*}}{\sigma s}\right) - T}{1-T} \tag{57}$$

Recall that $T = \Phi\left(\frac{1}{\sigma s}\left(\log\left(\frac{(\sigma-1)\underline{m}}{\underline{m}+\sigma-\underline{m}\sigma}\right) - \tilde{\mu}\right)\right)$ and $\tilde{\mu}_{x*} = \tilde{\mu} - \sigma \log\left(\frac{\sigma-1}{\sigma}\underline{m}\right)$. So, once again, conditional on observables $\underline{m}$, $\tilde{\mu}$, s and σ, the proportion of firms that are smaller or equal to x_c is independent of γ in both the market and the planner's distributions.

B Supplementary Online Appendix

Section B.1 proves some further corollaries of Proposition 1 beyond those given in Section 3. Section B.2 expands on Appendix A.3 with further properties of the CREMR demand function and its demand manifold. Section B.3 uses the Kullback-Leibler Divergence to compare the goodness of fit for both sales and markups of Indian firms of different assumptions about the productivity distribution and demand. Section B.4 explores how the relative performance of different specifications, especially the choice of Pareto versus lognormal productivity, is affected by truncating the sample. Section B.5 shows that the results are robust to an alternative distance measure, the QQ estimator. Finally, Section B.6 shows that similar results are obtained with a different data set of exports by French firms to Germany.

B.1 Other Implications of Proposition 1

The proof of each of these corollaries to Proposition 1 proceeds in the same way. Given two distributions $y \sim G(y)$ and $z \sim F(z)$, we first solve for $y = G^{-1}(F(z))$, and then solve the resulting differential equation to derive the implied demand function.

B.1.1 Self-Reflection of Productivity and Output

We first explore the conditions under which output follows the same distribution as productivity. Proposition 1 implies that a necessary and sufficient condition for this form of self-reflection is that productivity is a simple power function of output: $\varphi = \varphi_0 x^E$. Replacing φ by $r'(x)^{-1}$ as before yields a new differential equation in $r(x)$, the solution to which is:

$$p(x) = \frac{1}{x}(\alpha + \beta x^{\frac{\sigma-1}{\sigma}}) \tag{58}$$

This demand function plays the same role with respect to firm output as the CREMR demand function does with respect to firm sales (recall (6)). It is necessary and sufficient for a constant elasticity of marginal revenue with respect to output, equal to $E = \frac{1}{\sigma}$. Hence

we call it "CEMR" for "Constant (Output) Elasticity of Marginal Revenue."[26]

Unlike CREMR, there are some precedents for the CEMR class. It has the same functional form, except with prices and quantities reversed, as the direct PIGL ("Price-Independent Generalized Linearity") class of Muellbauer (1975).[27] In particular, the limiting case where σ approaches one is the inverse translog demand function of Christensen et al. (1975). However, except for the CES (the special case when $\alpha = 0$), CEMR demands bear little resemblance to commonly-used demand functions.[28]

When the common distribution of productivity and output is a Pareto, we can immediately state a further corollary of Proposition 1:

Corollary 4. *Given Assumption 1, any two of the following imply the third: (A) The distribution of firm productivity is Pareto:* $G(\varphi) = 1 - \underline{\varphi}^k \varphi^{-k}$*; (B) The distribution of firm output is Pareto:* $F(x) = 1 - \underline{x}^m x^{-m}$*; (C) The demand function belongs to the CEMR family in* (58)*; where the parameters are related as follows:*

$$\sigma = \frac{k}{m} \quad \Leftrightarrow \quad m = \frac{k}{\sigma} \qquad \textit{and} \qquad \beta = \frac{k}{k-m} \frac{\underline{x}^{\frac{m}{k}}}{\underline{\varphi}} \quad \Leftrightarrow \quad \underline{x} = \left(\beta \, \frac{\sigma - 1}{\sigma} \, \underline{\varphi}\right)^{\sigma} \tag{59}$$

A similar result holds if firm productivities have a lognormal distribution, though, as in the CREMR case of Corollary 2, we have to allow for the possibility that the distribution is left-truncated, as the value of output for the smallest firm may be strictly positive.

Corollary 5. *Given Assumption 1, any two of the following imply the third: (A) The distribution of firm productivity is truncated lognormal:* $G(\varphi) = \frac{\Phi((\log\varphi - \mu)/s) - T}{1-T}$*; (B) The distribution of firm output is truncated lognormal:* $F(r) = \frac{\Phi((\log x - \mu'')/s'') - T}{1-T}$*; (C) The demand*

[26] "CEMR" rhymes with "seemer," just as "CREMR" rhymes with "dreamer."

[27] For this reason, Mrázová and Neary (2017) called it the "inverse PIGL" class of demand functions.

[28] As shown by Mrázová and Neary (2017), the CEMR demand manifold implies a linear relationship between the convexity and elasticity of demand, passing through the Cobb-Douglas point $(\varepsilon, \rho) = (1, 2)$: $\rho = 2 - \frac{\varepsilon - 1}{\sigma}$. The manifold for the inverse translog special case ($\sigma \to 1$) coincides with the SM locus in Figure 2(b). For smaller firms when demand is subconvex, CEMR demands are qualitatively similar to CREMR, except that they are somewhat more elastic: the CEMR manifold can be written as $\varepsilon = (2-\rho)\sigma + 1$, while for high ε (so $\varepsilon - 1 \approx \varepsilon$) the CREMR manifold is approximately $\varepsilon = (2 - \rho)(\sigma - 1) + 1$.

function belongs to the CEMR family in (58)*; where the parameters are related as follows:*

$$\sigma = \frac{s''}{s} \quad \Leftrightarrow \quad s'' = \sigma s \tag{60}$$

$$\beta = \frac{s''}{s''-s} \exp\left(\frac{s}{s''}\mu'' - \mu\right) \quad \Leftrightarrow \quad \mu'' = \sigma\left(\mu + \log\left(\beta\frac{\sigma-1}{\sigma}\right)\right) \tag{61}$$

$$\underline{x} = \left(\underline{\varphi} \exp\left(\frac{s}{s''}\mu'' - \mu\right)\right)^{\frac{s''}{s}} = \left(\beta\frac{\sigma-1}{\sigma}\underline{\varphi}\right)^{\sigma} \tag{62}$$

$$T = \Phi((\log \underline{\varphi} - \mu)/s) = \Phi\left((\log \underline{x} - \mu'')/s''\right) \tag{63}$$

B.1.2 Self-Reflection of Output and Sales

A final self-reflection corollary of Proposition 1 relates to the case where output and sales follow the same distribution. This requires that the elasticity of one with respect to the other is constant, which implies that the demand function must be a CES.[29] Formally:

Corollary 6. ***Pareto Output and Sales Revenue:*** *Any two of the following statements imply the third: 1. The distribution of firm output x is a member of the generalized power function class; 2. The distribution of firm sales revenue r is the same member of the generalized power function class; 3. The demand function is CES: $p(x) = \beta x^{-\frac{1}{\sigma}}$, where $\beta = x_0^{-\frac{1}{E}}$ and $\sigma = \frac{E}{E-1}$.*

In the Pareto case, the sufficiency part of this result is familiar from the large literature on the Melitz model with CES demands: it is implicit in Chaney (2008) for example. The necessity part, taken together with earlier results, shows that it is not possible for all three firm attributes, productivity, sales and revenue, to have the same distribution from the generalized power class under any demand system other than the CES. Corollary 6 follows immediately from previous results when productivities themselves have a generalized power function distribution, since the only demand function which is a member of both the CEMR

[29] Suppose that $x = x_0 r(x)^E$. Recalling that $r(x) = xp(x)$, it follows immediately that the demand function must take the CES form.

and CREMR families is the CES itself. However, it is much more general than that, since it does not require any assumption about the underlying distribution of productivities. It is an example of a corollary to Proposition 1 which relates two endogenous firm outcomes rather than an exogenous and an endogenous one.

B.2 Further Properties of CREMR Demand Functions

To establish conditions for demand to be superconvex, we solve for the points of intersection between the demand manifold and the CES locus, the boundary between the sub- and superconvex regions. From Mrázová and Neary (2017), the expression for the CES locus is: $\rho = \frac{\varepsilon+1}{\varepsilon}$. Eliminating ρ using the CREMR demand manifold (21) and factorizing gives:

$$\rho - \frac{\varepsilon+1}{\varepsilon} = -\frac{(\varepsilon-\sigma)(\varepsilon-1)}{(\sigma-1)\varepsilon} = 0 \tag{64}$$

Given $1 < \sigma \leq \infty$, this expression is zero, and so every CREMR manifold intersects the CES locus, at two points. One is at $\{\varepsilon, \rho\} = \{1, 2\}$, implying that all CREMR demand manifolds must pass through the Cobb-Douglas point. The other is at $\{\varepsilon, \rho\} = \{\sigma, 1 + \frac{1}{\sigma}\}$. Hence every CREMR demand manifold lies strictly within the superconvex region (where $\rho > \frac{\varepsilon+1}{\varepsilon}$) for $\sigma > \varepsilon > 1$, and strictly within the subconvex region for $\varepsilon > \sigma$. The condition for superconvexity, $\varepsilon \leq \sigma$, can be reexpressed in terms of γ by using the fact that the elasticity of demand is $\varepsilon = \frac{x-\gamma}{x-\gamma\sigma}\sigma$. Substituting and recalling that σ must be strictly greater than one, we find that CREMR demands are superconvex if and only if $\gamma \leq 0$. As with many other demand manifolds considered in Mrázová and Neary (2017), this implies that, for a given value of σ, the demand manifold has two branches, one in the superconvex region corresponding to negative values of γ, and the other in the subconvex region corresponding to positive values of γ. Along each branch, the equilibrium point converges towards the CES locus as output rises without bound, as shown by the arrows in Figure 2.

Similarly, to establish conditions for profits to be supermodular, we solve for the points

of intersection between the demand manifold and the SM locus, the boundary between the sub- and supermodular regions. From Mrázová and Neary (2017), the expression for the SM locus is: $\rho = 3 - \varepsilon$. Eliminating ρ using the CREMR demand manifold and factorizing gives:

$$\rho + \varepsilon - 3 = \frac{((\sigma - 2)\varepsilon + 1)(\varepsilon - 1)}{(\sigma - 1)\varepsilon} = 0 \tag{65}$$

Once again, this expression is zero at two points: the Cobb-Douglas point $\{\varepsilon, \rho\} = \{1, 2\}$, and the point $\{\varepsilon, \rho\} = \{\frac{1}{2-\sigma}, \frac{5-3\sigma}{2-\sigma}\}$. The latter is in the admissible region only for $\sigma < 2$. Hence for $\sigma \geq 2$, the CREMR demand manifold is always in the supermodular region.

B.3 Fitting Sales and Markup Distributions

Section 6.2 in the text focused on how different assumptions compare in predicting the distribution of markups. Here we supplement this by showing in addition how they compare in predicting the distribution of sales. To compare the "goodness of fit" of different models, we use the Kullback-Leibler Divergence (denoted "KLD" hereafter), introduced by Kullback and Leibler (1951). This measures the divergence of the predicted distribution from the actual one, and is asymptotically equivalent to maximum likelihood.[30] It equals the information loss from using the theory rather than knowing the true distribution. Whereas information scientists typically present KLD values in "bits" (log to base 2) or "nats" (log to base e), units with little intuitive appeal in economics, we present its values normalized by the value implied by a uniform distribution. This is an uninformative prior in the spirit of the Laplace principle of insufficient reason; it is analogous to the "dartboard" approach to benchmarking the geographic concentration of manufacturing industry of Ellison and Glaeser (1997), or the "balls and bins" approach to benchmarking the world trade matrix of Armenter and Koren (2014). The value of the KLD is unbounded, but a specification that gave a value greater

[30] The KLD weights the log of the ratio of the estimated density to the empirical density by the empirical density itself. Many alternative weighting schemes have been proposed, such as Exponential Tilting, which weights the ratio of the empirical density to the estimated density by the estimated density. (See Nevo (2002) for further discussion).

than that implied by a uniform distribution would be an unsatisfactory explanation of the data. Appendix B.5 shows that an alternative criterion for choosing between distributions, the QQ estimator, gives qualitatively similar results.

	CREMR	Translog	LES	Linear
	A. Sales			
Pareto	0.2253	0.1028	0.1837	0.1837
Lognormal	0.0140	0.5825	0.7266	0.7266
	B. Markups			
Pareto	0.1851	0.2205	0.2191	0.2512
Lognormal	0.1863	0.2228	0.2083	0.2075

Table 5: KLD for Indian Sales and Markups Compared with Predictions from Selected Productivity Distributions and Demand Functions

Note: Each KLD value measures the divergence of the predicted from the empirical distribution. A value of zero indicates no divergence, a value of one a divergence as great as a uniform distribution.

The minimized KLD values for each specification are given in Table 5 and illustrated in Figure 8. The rankings of different specifications for sales are very different in the Pareto and lognormal cases. Conditional on a Pareto distribution of productivities, CREMR demands give the worst fit to sales, with translog demands performing best, and linear-LES intermediate between the others. However, the differences between the KLD values for these specifications are much less than those conditional on lognormal productivities. In this case CREMR does best, with translog performing much less well and linear-LES worst of all.

As for the results for markups, these imply exactly the same ranking of different specifications as the estimates given in Table 2 in Section 6.2, despite the different methodologies used (using individual observations and minimizing the AIC rather than using data grouped in bins and minimizing the KLD as here). Once again, CREMR demands clearly do best, irrespective of the assumed distribution, with translog and LES performing at the same level, and linear doing better under Pareto assumptions but less well in the lognormal case. This reinforces the conclusion drawn in the text that the choice between Pareto and lognormal

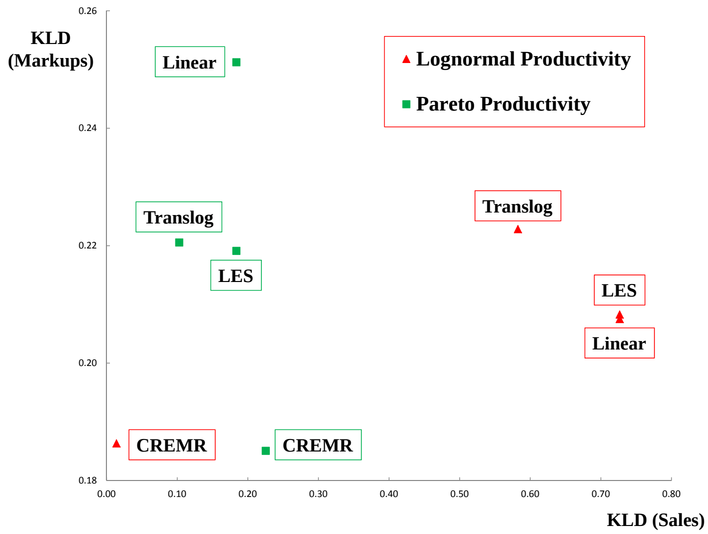


Figure 8: KLD for Indian Sales and Markups

Note: Data from Table 5.

	CREMR + LN	CREMR + P	TLog + P	Lin + P	TLog + LN	Lin + LN
CREMR + LN	–	0%	0%	0%	0%	0%
CREMR + P	100%	–	0%	0%	0%	0%
TLog + P	100%	100%	–	0%	0%	0%
Lin + P	100%	100%	100%	–	0%	0%
TLog + LN	100%	100%	100%	100%	–	0%
Lin + LN	100%	100%	100%	100%	100%	–

Table 6: Bootstrapped Robustness of the KLD Ranking: Indian Sales

Note: See text for explanation. "LN" denotes lognormal, "P" denotes Pareto.

distributions is less important than the choice between CREMR and other demands.

To assess whether the KLD values in Table 5 are significantly different from one another, we use a bootstrapping approach. We construct one thousand samples of the same size as the data (i.e., 2,457 observations), by sampling with replacement from the original data. For each sample, we then compute the KLD value for each of the six models. Tables 6 and 7 gives the results for Indian sales and markup data respectively. Each entry in the table is the proportion of samples in which the combination in the relevant column gives a higher value of the KLD than that in the relevant row. All the values are equal to or very close to

	CREMR + P	CREMR + LN	Lin + LN	LES + LN	LES + P	TLog + P	TLog + LN	Lin + P
CREMR + P	–	2%	0%	0%	0%	0%	0%	0%
CREMR + LN	98%	–	0%	0%	0%	0%	0%	0%
Lin + LN	100%	100%	–	0%	0%	0%	0%	0%
LES + LN	100%	100%	100%	–	0%	0%	0%	0%
LES + P	100%	100%	100%	100%	–	16%	6%	0%
TLog + P	100%	100%	100%	100%	84%	–	0%	0%
TLog + LN	100%	100%	100%	100%	94%	100%	–	0%
Lin + P	100%	100%	100%	100%	100%	100%	100%	–

Table 7: Bootstrapped Robustness of the KLD Ranking: Indian Markups

Note: See text for explanation.

100%, which confirms that the results in Table 5 are robust.

B.4 Robustness to Truncation

The results for Indian sales data in the preceding sub-section are broadly similar to those with French sales data in Appendix B.6 below, except for the case of CREMR demands combined with Pareto productivity: this gives a good fit with French data but performs less well with Indian data. One possible explanation for this is that the French data relate to exports, whereas the Indian data are for total domestic production. Presumptively, smaller firms have been selected out of the French data, so we might expect the Pareto assumption to be more appropriate. To throw light on this issue, we explore the robustness of the Indian results to left-truncating the data: specifically, we repeat a number of the comparisons between different specifications for the Indian sales distribution dropping one observation at a time.

Figure 9 compares the KLD for the Pareto and lognormal, conditional on CREMR demands, starting on the left-hand side with all observations (so the values are the same as in Figure 8) and successively dropping up to 809 observations one at a time.[31] Although the curves are not precisely monotonic, the broad picture is clear: conditional on CREMR demands, Pareto does better and lognormal does worse as more and more observations are

[31]Each KLD value is normalized by the value of the KLD for a uniform distribution corresponding to the number of observations used to calculate it; i.e., excluding the observations dropped. Alternative approaches would make very little difference however, as the KLD value for the uniform varies very little, from 3.940 with no observations dropped to 3.560 with 809 observations dropped.

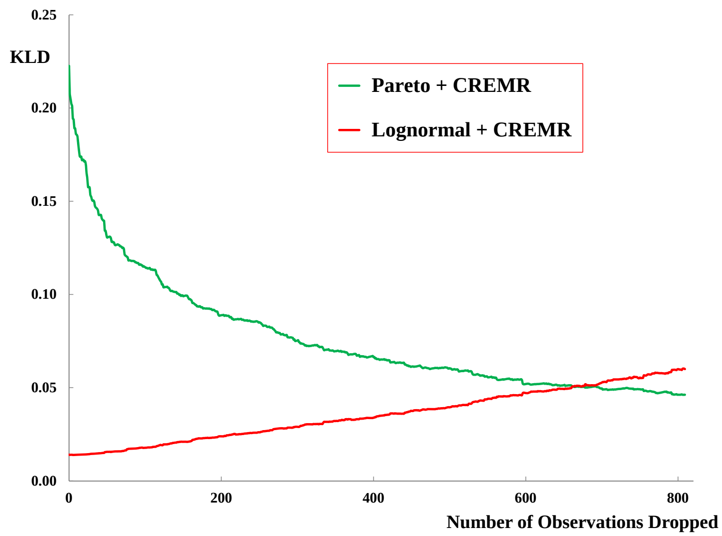


Figure 9: CREMR vs. CREMR: KLD for Indian Sales

dropped. The Pareto specification dominates when we drop 663 or more observations: these account for 27% of all firm-product observations, but only 1.2% of total sales.

Figure 10 shows that a similar pattern emerges when we compare the performance of different demand functions in explaining the sales distribution, conditional on a Pareto distribution for productivity. (Note that the horizontal scale differs from that in Figure 9.) In this case, the CREMR specification overtakes the linear one when we drop 11 or more observations, which account for 0.44% of all firm-product observations, and only 0.0002% of sales. As for the translog, CREMR overtakes it when we drop 118 or more observations, which account for 4.80% of observations, and 0.03% of sales.

These findings confirm that the combination of CREMR demand and Pareto productivities fits the sales data relatively better when the smallest observations are dropped. They also make precise the pattern observed in many datasets, whereby the Pareto assumption outperforms the lognormal in the right tail of the sales distribution. For example, Figure 9 shows that the relevant region in the right tail begins at exactly 663 observations.

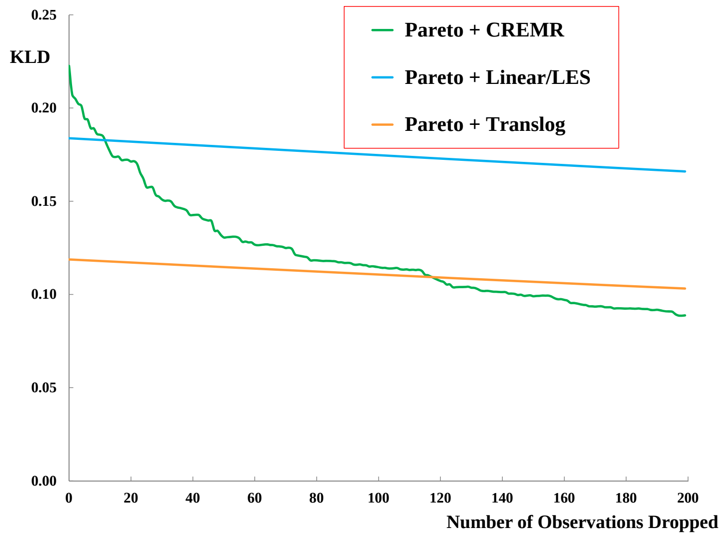


Figure 10: CREMR vs. The Rest, Given Pareto: KLD for Indian Sales

B.5 Robustness to Divergence Criterion: The QQ Estimator

To check the robustness of our results, we consider an alternative criterion to the KLD for comparing predicted and actual distributions. Here we consider the QQ estimator, developed by Kratz and Resnick (1996), and previously used by Head et al. (2014) and Nigai (2017). This estimator does not have the same desirable theoretical properties as the KLD, in particular it is not asymptotically equivalent to maximum likelihood, but it has a simple interpretation. It equals the parameter vector $\boldsymbol{\theta}^*$ that minimizes the sum of the squared deviations of the quantiles of the predicted distribution from those of the actual distribution:

$$QQ(\tilde{F} \mid\mid F(\cdot;\boldsymbol{\theta})) = \sum_{i=1}^{n} \left(\log \tilde{q}_i - \log q_i(\boldsymbol{\theta})\right)^2 \tag{66}$$

where $\tilde{q}_i = \tilde{F}^{-1}(i/n)$ is the i'th quantile observed in the data, while $q_i(\boldsymbol{\theta}) = F^{-1}(i/n;\boldsymbol{\theta})$ is the i'th quantile predicted by the theory.

To implement the QQ estimator we need analytic expressions for the quantiles of the sales and markup distributions under each of the eight combinations of assumptions about

Demand	Pareto $\mathcal{P}(\underline{\varphi}, k)$	lognormal $\mathcal{LN}(\mu, s)$
CREMR	$\underline{r}\,(1-y)^{-\frac{\sigma-1}{k}}$	$exp\left(\mu + s \cdot \Phi^{-1}(y)\right)$
Translog	$\eta \cdot \left(\mathcal{W}\left(e \cdot (1-y)^{-\frac{1}{k}}\right) - 1\right)$	$\eta \cdot \left(\mathcal{W}\left(exp\left(\frac{\gamma}{\eta} + 1 + s \cdot \Phi^{-1}(y) + \mu\right)\right) - 1\right)$
Linear/LES	$\overline{r}\left(1 - (1-y)^{\frac{1}{2k}}\right)$	$\overline{r} - \frac{exp\left(-2\left(\mu + s \cdot \Phi^{-1}(y)\right)\right)}{4\beta}$

Table 8: Quantiles for Sales

Note: $\Phi[z]$: c.d.f. of a standard normal; $\mathcal{W}$: The Lambert function.

Demand	Pareto $\mathcal{P}(\underline{\varphi}, k)$	lognormal $\mathcal{LN}(\mu, s)$
CREMR	$\frac{\overline{m} \cdot (1-y)^{-\frac{1}{k}}}{\overline{m} - 1 + (1-y)^{-\frac{1}{k}}}$	$\overline{m}\left(1 + exp\left(-\mu - s \cdot \zeta\left(y; \frac{1}{s} \cdot \left(\log\left(\frac{1}{\overline{m}-1}\right) - \mu\right)\right)\right)\right)^{-1}$
Translog	$\mathcal{W}\left(e \cdot (1-y)^{-\frac{1}{k}}\right)$	$\mathcal{W}\left(exp\left(s \cdot \zeta\left(y; -\frac{1}{s} \cdot \left(\frac{\gamma}{\eta} + \mu\right)\right) + \frac{\gamma}{\eta} + \mu + 1\right)\right)$
Linear	$\frac{1}{2} + \frac{1}{2} \cdot (1-y)^{-\frac{1}{k}}$	$\frac{1}{2} + \frac{\alpha}{2} exp\left(\mu + s \cdot \zeta\left(y; -\frac{1}{s} \cdot (\log(\alpha) + \mu)\right)\right)$
LES	$(1-y)^{-\frac{1}{2 \cdot k}}$	$\sqrt{\frac{\delta}{\gamma}} \cdot exp\left(\frac{\mu}{2} + \frac{s}{2} \cdot \zeta\left(y, \frac{1}{s} \cdot \left(\log\left(\frac{\gamma}{\delta}\right) - \mu\right)\right)\right)$

Table 9: Quantiles for Markups

Note: $\zeta[y; z] \equiv \Phi^{-1}\left((1 - \Phi(z)) \cdot y + \Phi(z)\right)$.

demand and the distribution of productivity we consider. These are given in Tables 8 and 9. We set the number of quantiles n equal to 100. The resulting values of the QQ estimator for Indian sales and markups are given in Table 10, and they are illustrated in Figure 11.

Comparing Table 10 and Figure 11 with Table 5 and Figure 8 in Appendix B.3 respectively, it is evident that the results based on the QQ estimator are qualitatively very similar to those for the KLD. In particular, the Pareto assumption gives a better fit for sales than for markups, except in the CREMR case; while the lognormal assumption tends to give a better fit for markups than for sales. Comparing different demand functions, CREMR demands give a better fit to the markup distribution than any other demands, irrespective of which productivity distribution is assumed. As for sales, the results differ between the Pareto and lognormal cases. Conditional on lognormal, CREMR again performs much better, whereas, conditional on Pareto, it performs least well, with the translog doing best. The only quali-

tative difference between the results using the two criteria is that with the QQ estimator the translog does somewhat better than the LES in fitting the markup distribution. Overall, we can conclude that the rankings given in Section 6.1 are not unduly sensitive to our choice of criterion for comparing actual and predicted distributions.

	CREMR	Translog	LES	Linear
	A. Sales			
Pareto	58.939	12.693	24.484	24.484
Lognormal	3.078	116.918	133.274	133.274
	B. Markups			
Pareto	0.113	0.978	1.133	3.606
Lognormal	0.110	0.990	0.340	0.325

Table 10: QQ Estimator for Indian Sales and Markups

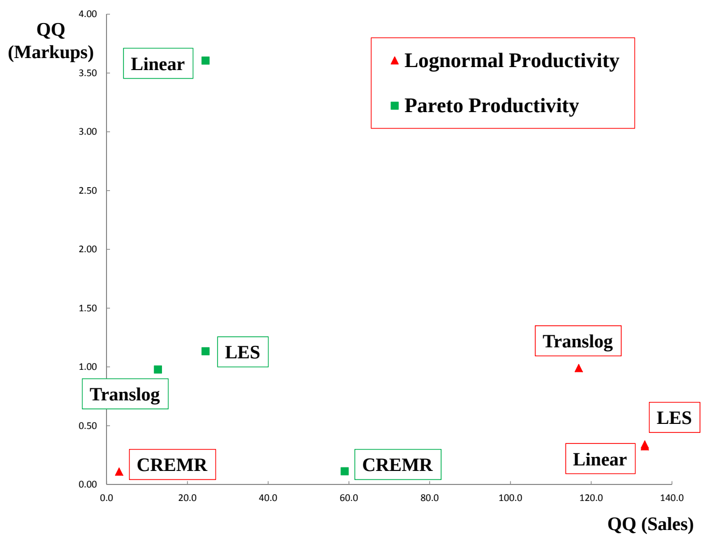


Figure 11: QQ Estimator for Indian Sales and Markups

B.6 French Exports to Germany

The Indian data used in Section 6 have the great advantage that they give both sales and markups for all firms. This is important, for example, in allowing us to discriminate between CES and CREMR, whose implications for sales are observationally equivalent. However, relative to many data sets used in recent trade applications, they refer to total sales rather than exports and they cover a relatively small number of firms. Hence it is useful to repeat the analysis on a more conventional data set on export sales, even if this does not give information on markups. We do this in this section, using data on the universe of French exports to Germany in 2005, drawn from the same source as that used by Head et al. (2014).[32]

	CREMR/CES	Translog	Linear and LES
Pareto	0.0012	0.3819	0.4711
Lognormal	0.0001	0.7315	0.8314

Table 11: KLD for French Exports Compared with Predictions from Selected Demand Functions and Productivity Distributions

As with the Indian data in the text, we use the KLD as the criterion to determine how well different assumptions fit the data. Table 11 gives the values of the KLD measuring the divergence from the empirical sales distribution from the distributions implied by CREMR/CES, translog and linear demand functions combined with either Pareto or lognormal productivities. These distributions are calculated by combining the relevant productivity distribution with the relationships between productivity and sales given in Table 1. (Recall from that table that the linear and LES specifications are observationally equivalent.) Each entry in the table is the value of the KLD that measures the information loss when the combination of assumptions indicated by the row and column is used to explain the observed distribution of sales. (As with the Indian data in Section 6, the data are normalized by the value of the KLD for a uniform distribution, which for this data set is 6.8082.)

[32]The data set contains 161,191 firm-product observations on export sales by 27,550 firms: 5.85 products per firm.

Turning to the results in Table 11, the values of the minimized KLD show that, conditional on CREMR or CES demands, the lognormal provides a better overall fit than the Pareto: 0.0001 as opposed to 0.0012. However, the difference between distributions turns out to be much less significant than those between different specifications of demand. The KLD values for the translog and linear/LES specifications are much higher than for the CREMR case, as shown in the third and fourth columns of Table 11, with the Pareto now preferred to the lognormal. The overwhelming conclusion from these results is that, if we want to fit the distribution of sales in this data set, then the choice between Pareto and lognormal distributions is less important than the choice between CREMR and other demands. This is broadly in line with the results for Indian sales data in Section 6, especially when we exclude the smallest firms as in Section B.4.

Table 12 repeats for French exports data the bootstrapping comparisons presented in Tables 6 and 7 for Indian sales and markup data respectively. It is clear that the comparisons between different values of the KLD for the French data shown in Table 11 are just as robust as those for the Indian data shown in Table 5 and Figure 8 in the text.

	CREMR + LN	CREMR + P	TLog + P	Lin + P	TLog + LN	Lin + LN
CREMR + LN	–	0%	0%	0%	0%	0%
CREMR + P	100%	–	0%	0%	0%	0%
TLog + P	100%	100%	–	0%	0%	0%
Lin + P	100%	100%	100%	–	0%	0%
TLog + LN	100%	100%	100%	100%	–	0.3%
Lin + LN	100%	100%	100%	100%	99.7%	–

Table 12: Bootstrapped Robustness of the KLD Ranking: French Sales

Note: See text for explanation.

References

Antràs, P. and D. Chor (2013): "Organizing the Global Value Chain," *Econometrica*, 81, 2127–2204.

Antràs, P. and E. Helpman (2004): "Global Sourcing," *Journal of Political Economy*, 112, 552–580.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2018): "The Elusive Pro-Competitive Effects of Trade," *Review of Economic Studies*, 86, 46–80.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

Armenter, R. and M. Koren (2014): "A Balls-and-Bins Model of Trade," *American Economic Review*, 104, 2127–2151.

Atkeson, A. and A. Burstein (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98, 1998–2031.

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820.

Bas, M., T. Mayer, and M. Thoenig (2017): "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity," *Journal of International Economics*, 108, 1–19.

Bee, M. and S. Schiavo (2018): "Powerless: Gains from Trade when Firm Productivity is not Pareto Distributed," *Review of World Economics*, 154, 15–45.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014): "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 122, 507–553.

Behrens, K., G. Mion, Y. Murata, and J. Suedekum (2020): "Quantifying the Gap Between Equilibrium and Optimum under Monopolistic Competition," *Quarterly Journal of Economics*, 135, 2299–2360.

Bernard, A. B., S. J. Redding, and P. K. Schott (2011): "Multi-Product Firms and Trade Liberalization," *Quarterly Journal of Economics*, 126, 1271–1318.

Bertoletti, P. and P. Epifani (2014): "Monopolistic Competition: CES Redux?" *Journal of International Economics*, 93, 227–238.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2020): "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data," NBER Working Paper No. 27002.

Bustos, P. (2011): "Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American Economic Review*, 101, 304–340.

Cabral, L. M. and J. Mata (2003): "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 93, 1075–1090.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–1721.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1975): "Transcendental Logarithmic Utility Functions," *American Economic Review*, 65, 367–383.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, Markups and Trade Reform," *Econometrica*, 84, 445–510.

Deaton, A. and J. Muellbauer (1980): "An Almost Ideal Demand System," *American Economic Review*, 70, 312–326.

Dhingra, S. and J. Morrow (2019): "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," *Journal of Political Economy*, 127, 196–232.

di Giovanni, J. and A. A. Levchenko (2012): "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies," *Journal of Political Economy*, 120, 1083–1132.

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

Eaton, J. and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

Edmond, C., V. Midrigan, and D. Y. Xu (2015): "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 105, 3183–3221.

Eeckhout, J. (2004): "Gibrat's Law for (All) Cities," *American Economic Review*, 94, 1429–1451.

Ellison, G. and E. L. Glaeser (1997): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105, 889–927.

Epifani, P. and G. Gancia (2011): "Trade, Markup Heterogeneity and Misallocations," *Journal of International Economics*, 83, 1–13.

Fabinger, M. and E. G. Weyl (2012): "Pass-Through and Demand Forms," working paper, University of Chicago.

Fally, T. (2018): "Integrability and Generalized Separability," NBER Working Paper No. 25025.

Feenstra, R. and H. L. Kee (2008): "Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity," *Journal of International Economics*, 74, 500–518.

Feenstra, R. C. (2018): "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," *Journal of International Economics*, 110, 16–27.

Feenstra, R. C. and D. E. Weinstein (2017): "Globalization, Markups, and U.S. Welfare," *Journal of Political Economy*, 125, 1040–1074.

Gabaix, X. (1999): "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 114, 739–767.

——— (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–294.

——— (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

Gibrat, R. (1931): *Les inégalités économiques: applications d'une loi nouvelle, la loi de l'effet proportionnel [Economic Inequalities: Applications of a New Law, the Law of Proportionate Effect]*, Librairie du Recueil Sirey.

Hallak, J. C. and J. Sivadasan (2013): "Product and Process Productivity: Implications for Quality Choice and Conditional Exporter Premia," *Journal of International Economics*, 91, 53–67.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013): "Who Creates Jobs? Small versus Large versus Young," *Review of Economics and Statistics*, 95, 347–361.

Haltiwanger, J., R. Kulick, and C. Syverson (2018): "Misallocation Measures: The Distortion that Ate the Residual," NBER Working Paper No. 24199.

Harrigan, J. and A. Reshef (2015): "Skill Biased Heterogeneous Firms, Trade Liberalization, and the Skill Premium," *Canadian Journal of Economics*, 48, 1024–1066.

Hayashi, F. (2000): *Econometrics*, Princeton: Princeton University Press.

Head, K., T. Mayer, and M. Thoenig (2014): "Welfare and Trade Without Pareto," *American Economic Review, Papers and Proceedings*, 104, 310–316.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004): "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, 94, 300–316.

Holmes, T. J. and J. J. Stevens (2014): "An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade," *Journal of Political Economy*, 122, 369–421.

Hopenhayn, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60, 1127–1150.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

Ijiri, Y. and H. A. Simon (1974): "Interpretations of Departures from the Pareto Curve Firm-Size Distributions," *Journal of Political Economy*, 82, 315–331.

Johnson, N. (1949): "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36, 149–176.

Jovanovic, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50, 649–670.

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 27, 1241–1277.

Klenow, P. J. and J. L. Willis (2016): "Real Rigidities and Nominal Price Changes," *Economica*, 83, 443–472.

Kratz, M. and S. I. Resnick (1996): "The QQ-estimator and heavy tails," *Stochastic Models*, 12, 699–724.

Krusell, P. and A. A. Smith, Jr. (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867–896.

Kullback, S. and R. A. Leibler (1951): "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79–86.

Lamorgese, A., A. Linarello, and F. Warzynski (2014): "Free Trade Agreements and Firm-Product Markups in Chilean Manufacturing," Tech. rep., Aarhus University.

Luttmer, E. G. (2010): "Models of Growth and Firm Heterogeneity," *Annual Reviews of Economics*, 2, 547–576.

Matsuyama, K. and P. Ushchev (2017): "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems," CEPR Discussion Paper No. 12210.

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

Mead, R. (1965): "A Generalised Logit-Normal Distribution," *Biometrics*, 21, 721–732.

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

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

Melitz, M. J. and S. J. Redding (2015): "New Trade Models, New Welfare Implications," *American Economic Review*, 105, 1105–46.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.

Mrázová, M. and J. P. Neary (2017): "Not So Demanding: Demand Structure and Firm Behavior," *American Economic Review*, 107, 3835–3874.

——— (2019): "Selection Effects with Heterogeneous Firms," *Journal of the European Economic Association*, 17, 1294–1334.

Muellbauer, J. (1975): "Aggregation, Income Distribution and Consumer Demand," *Review of Economic Studies*, 42, 525–543.

Nevo, A. (2002): "Sample Selection and Information-Theoretic Alternatives to GMM," *Journal of Econometrics*, 107, 149–157.

Nigai, S. (2017): "A Tale of Two Tails: Productivity Distribution and the Gains from Trade," *Journal of International Economics*, 104, 44–62.

Nocco, A., G. I. P. Ottaviano, and M. Salto (2014): "Monopolistic Competition and Optimum Product Selection," *American Economic Review*, 104, 304–09.

Parenti, M., P. Ushchev, and J.-F. Thisse (2017): "Toward a Theory of Monopolistic Ccompetition," *Journal of Economic Theory*, 167, 86–115.

Pollak, R. A. (1972): "Generalized Separability," *Econometrica*, 40, 431–453.

Simonovska, I. (2015): "Income Differences and Prices of Tradables: Insights from an Online Retailer," *Review of Economic Studies*, 82, 1612–1656.

Stiglitz, J. E. (1969): "Distribution of Income and Wealth Among Individuals," *Econometrica*, 37, 382–397.

Sutton, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35, 40–59.

Tintelnot, F. (2017): "Global Production with Export Platforms," *Quarterly Journal of Economics*, 132, 157–209.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012): "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," *Econometrica*, 80, 2765–2784.