SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS*

Monika Mrázová[†]
University of Geneva,
and CEPR

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

Mathieu Parenti[§]
ULB and CEPR

December 19, 2015

Revised June 28, 2016

*Earlier versions were circulated 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, and to Abi Adams, Jonathan Dingel, Peter Egger, Basile Grassi, Joe Hirschberg, Rosa Matzkin, Kevin Roberts, Gonzague Vannoorenberghe, Frank Windmeijer and Tianhao Wu for helpful comments. 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; e-mail: monika.mrazova@unige.ch.

[‡]Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK; e-mail: peter.neary@economics.ox.ac.uk.

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

Abstract

We derive exact conditions relating the distributions of firm productivity, sales, output, and markups to the form of demand; in particular, for a large family (including Pareto, log-normal, and Fréchet), the distributions of productivity and output are the same if and only if demand is "CREMR" (Constant Revenue Elasticity of Marginal Revenue). We then use the Kullback-Leibler Divergence to quantify the information loss when a predicted distribution fails to match the actual one; and we find that, to explain sales and markups, the choice between Pareto and log-normal productivity distributions matters less than the choice between CREMR and other demands.

Keywords: CREMR Demands; Heterogeneous Firms; Kullback-Leibler Divergence; Log-Normal Distribution; Pareto Distribution.

JEL Classification: F15, F23, F12

1 Introduction

Models of monopolistic competition with heterogeneous firms have provided a fertile laboratory for studying a range of problems relating to the process of globalization. Much of this work to date has assumed special forms for preferences and technology, usually CES preferences on the demand side, and a Pareto distribution of firm productivity on the supply side. Well-known examples include Helpman, Melitz, and Yeaple (2004), Chaney (2008) and Arkolakis, Costinot, and Rodríguez-Clare (2012). More recently, a number of contributions has explored the implications of relaxing these assumptions. The implications of preferences other than CES have been considered by Melitz and Ottaviano (2008), Simonovska and Waugh (2011), Zhelobodko, Kokovin, Parenti, and Thisse (2012), Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012), Mrázová and Neary (2013), Bertoletti and Epifani (2014), and Parenti, Thisse, and Ushchev (2014), among others. Alternatives to the Pareto distribution have been less widely explored, though the pioneering work of Melitz (2003) avoided making explicit distributional assumptions, and Bee and Schiavo (2014) and Head, Mayer, and Thoenig (2014) consider the implications of a log-normal distribution.

Recent work has also drawn attention to the distribution of markups across firms. De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) and Lamorgese, Linarello, and Warzynski (2014) present suggestive evidence that markups exhibit an approximately log-normal distribution, but to date there is no model of industry equilibrium which would generate such patterns.

In this paper, we provide a general characterization of the problem of explaining the distribution of firm size and firm markups, given particular assumptions about the structure of demand and the distribution of firm productivities. We present two different kinds of results. On the one hand, we present exact conditions under which specific assumptions about the distribution of firm productivity are consistent with a particular form of the distribution of sales revenue, output, or markups. On the other hand, we use the Kullback-Leibler divergence from information theory to quantify the information loss when a predicted distribution fails to match the actual one. We show that applying this tool in the context of models of heterogeneous firms leads to new insights about the relationship between fundamentals and the size distribution of firms, and also provides a quantitative framework for gauging how well a given set of assumptions explain a given data set.

The first part of the paper presents exact characterizations of the links between the distributions of firm attributes, technology and preferences. We begin in Section 2 with two general propositions which characterize the form that very general distributions of firm characteristics and general models of firm behavior must take if they are to be mutually consistent. Sections 3 and 4 then apply these results to distributions of sales and markups respectively.[1] Among the results we derive is a characterization of the demand functions which are necessary and sufficient for productivity and sales to exhibit the same distribution from a wide family which includes Pareto, log-normal and Fréchet. We show that this property is implied by a new family of demands, a generalization of the CES, which we call "CREMR" for "Constant Revenue Elasticity of Marginal Revenue."[2] The CREMR class has many desirable properties but is very different from most of the non-CES demand systems used in applied economics. We also derive the distributions of markups which are implied by CREMR and other demand functions.

The second part of the paper addresses the question of how to proceed when the conditions for exact consistency between distributions, preferences and technology do not hold. Section 5 presents some tools from information theory which have a natural application in this context, explores their implications in the context of evaluating how "close" a predicted distribution comes to an actual one, and shows how these tools make it possible to quantify the cost of using the "wrong" assumptions about demand or technology to calibrate a hypothetical distribution of firm sales or markups. Section 6 illustrates how the tools can be applied to actual data sets. Finally, Section 7 concludes, while the Appendix contains proofs and more technical details.

[1] We use "sales" throughout to refer to sales revenue.

[2] "CREMR" rhymes with "dreamer".

2 Characterizing Links Between Distributions

The two central results of the paper link the distributions of two firm characteristics to a general specification of the relationship between them: we make no assumptions about whether either characteristic is exogenous or endogenous, nor about the details of the technological and demand constraints faced by firms which generate the relationship. All we assume is a hypothetical dataset of a continuum of firms, which reports for each firm i its characteristics $x(i)$ and $y(i)$, both of which are monotonically increasing functions of i.[3] Formally:

Assumption 1. $\{i, x(i), y(i)\} \in \left[\Omega \times (\underline{x}, \overline{x}) \times (\underline{y}, \overline{y})\right]$, *where Ω is the set of firms, with both $x(i)$ and $y(i)$ monotonically increasing functions of i.*

Examples of $x(i)$ and $y(i)$ include firm productivity, sales and output in most models of heterogeneous firms.

Our first result adapts a standard result in mathematical statistics to our context; it is closely related to Lemma 1 of Matzkin (2003).

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

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

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

(3) Firm behavior, given technology and demand, is such that: $x = v(y)$, $v'(y) > 0$;

provided the functions are related as follows:

(i) (1) and (3) imply (2) with $F(y) = G[v(y)]$ and $f(y) = g[v(y)]v'(y)$; similarly, (2) and (3) imply (1) with $G(x) = F[v^{-1}(x)]$ and $g(x) = f[v^{-1}(x)]\frac{d[v^{-1}(x)]}{dx}$.

[3] The assumption that they are increasing functions is without loss of generality. For example, if $x(i)$ is increasing and $y(i)$ is decreasing, Proposition 1 can easily be reformulated using the survival function of y. Monotonicity here is a property of theoretical models. In our empirical applications we do not need to assume that any measured firm characteristics are monotonic in i. We follow standard models of firm heterogeneity under monopolistic competition by considering a continuum of firms 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.

(ii) (1) and (2) imply (3) with $v(y) = G^{-1}[F(y)]$.

Part (i) of the proposition is a standard result on transformations of variables. Part (ii) is less standard, and requires Assumption 1: characteristics $x(i)$ and $y(i)$ must refer to the same firm and must be monotonically increasing in i.[4] The proof is in Appendix A. The importance of the result is that it allows us to characterize fully the conditions under which assumptions about distributions and about the functional forms that link them are mutually consistent. Part (ii) in particular provides an easy way of determining which specifications of firm behavior are consistent with particular assumptions about the distributions of firm characteristics. All that is required is to reverse-engineer the form of $v(y)$ implied by any pair of distributional assumptions.

Our next result shows how Proposition 1 is significantly strengthened when the distributions of the two firm characteristics share a common parametric structure.

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

(1) x *is distributed with CDF:* $G(x; \boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2} x^{\theta_2}\right)$, $G_x > 0$;

(2) y *is distributed with CDF:* $F(y; \boldsymbol{\theta'}) = G\left[h(y); \theta'\right] = H\left(\theta_0 + \frac{\theta'_1}{\theta'_2} h(y)^{\theta'_2}\right)$, $F_y > 0$

(3) $x = x_0 h(y)^E$;

provided the parameters are related as follows:

(i) (1) and (3) imply (2) with $\theta'_1 = E\theta_1 x_0^{\theta_2}$ *and* $\theta'_2 = E\theta_2$*; similarly, (2) and (3) imply (1) with* $\theta_1 = E^{-1}\theta'_1 x_0^{-E^{-1}\theta'_2}$ *and* $\theta_2 = E^{-1}\theta'_2$.

(ii) (1) and (2) imply (3) with $x_0 = \left(\frac{\theta_2}{\theta_1}\frac{\theta'_1}{\theta'_2}\right)^{\frac{1}{\theta_2}}$ *and* $E = \frac{\theta'_2}{\theta_2}$.

We call the functional form $H\left(\theta_0 + \frac{\theta_1}{\theta_2} x^{\theta_2}\right)$ in (1) a Generalized Power Function (henceforward "GPF") family of distributions. It is generalized in two respects: the elementary power function $\frac{\theta_1}{\theta_2} x^{\theta_2}$ is shifted by a constant θ_0, and is then subject to an arbitrary monotonic

[4] This implies that the Spearman rank correlation between x and y is one.

transformation $H(\cdot)$. As for the functional form in (2), it is the same as that in (1), except that the θ_1 and θ_2 parameters are different, and y is then subject to an arbitrary monotonic transformation $h(y)$. Both the $H(\cdot)$ and $h(\cdot)$ functions are completely general, and the components of the parameter vector $\boldsymbol{\theta} = \{\theta_0, \theta_1, \theta_2\}$ can take on any values, except in three respects: the derivative of H at every point must have the same sign as θ_1, from the strict monotonicity restriction on G: $H'\theta_1 > 0$ since $G_x = H'\theta_1 x^{\theta_2 - 1} > 0$; h must be monotonically increasing from the corresponding restriction on F: $h' > 0$ since $F_y = G_x h' > 0$; and θ_0 must be the same in both distributions, so both $G(x; \boldsymbol{\theta})$ and $F(y; \boldsymbol{\theta'})$ are two-parameter families of distributions.

The GPF family of distributions characterized by $H(\cdot)$ nests many of the most widely used in applied economics, including Pareto, truncated Pareto, log-normal, uniform, Fréchet, Gumbel, and Weibull. (See Appendix B for the proof of Proposition 2 and Appendix C for details of the GPF family.) Each choice of the $h(\cdot)$ function generates in turn a further family, such that the transformation $h(y)$ follows a distribution from the GPF family.[5] Proposition 2 shows that these families are intimately linked via a simple power function that expresses one of the two firm characteristics as a transformation of the other. In much of the paper we will concentrate on two special forms for the $h(\cdot)$ function. The identity transformation, $h(y) = y$, implies from Proposition 2 a property we call "self-reflection", since the distributions of x and y are the same. This case proves particularly useful when we consider distributions of firm sales and output in Section 3. The other case we consider in detail is the odds transformation, $h(y) = \frac{y}{1-y}$, where $0 \leq y \leq 1$. This implies a property we call "odds-reflection", since the distribution of y is an odds transformation of that of x. This case proves particularly useful when we consider distributions of firm markups in Section 4.

In the next two sections we give some examples of links between distributions and models of firm behavior implied by Propositions 1 and 2, with detailed derivations in Appendix F.

[5] Assuming that a transformation of a variable follows a standard distribution is a well-known method of generating new functional forms for distributions. See Johnson (1949), who attributes it to Edgeworth, and Jones (2015).

3 Backing Out Demands

The first set of applications of Proposition 2 apply part (ii) of the proposition: we ask what demand functions are consistent with assumed distributions of two different firm attributes. Moreover, following the existing literature, we ask when will we observe self-reflection, in the sense that the distributions of the two attributes are the same (though with different parameters of course). Figure 1 summarizes schematically the results of this section, which specify the demand functions that are necessary and sufficient for self-reflection between the distributions of any two of firm output x, sales revenue r, and productivity φ.

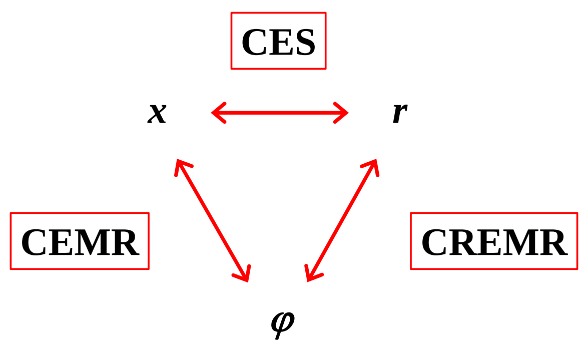


Figure 1: Links Between Firm Characteristics

3.1 Self-Reflection of Productivity and Sales

We begin in this sub-section by focusing on the two central attributes of productivity and sales revenue. We know from Helpman, Melitz, and Yeaple (2004) and Chaney (2008) that CES demands are sufficient to bridge the gap between two Pareto distributions; and we also know from Head, Mayer, and Thoenig (2014) that a log-normal distribution of productivity coupled with CES demands implies a log-normal distribution of sales. We want to establish the necessary conditions for these links, which in turn will tell us whether there are other demand systems that ensure an exact correspondence between the form of the productivity and sales distributions.

The answer to these questions is immediate from Proposition 2: if both productivity

φ and sales r follow the same distribution, which can be any member of the GPF family, including the Pareto and the log-normal, then they must be related by a power function: $\varphi = \varphi_0 r^E$. The implications of this for demand are immediate, provided we assume only that firms equate marginal cost to marginal revenue, so $\varphi = c^{-1} = (r')^{-1}$.[6] Combining these gives a simple differential equation in sales revenue, which is a function of output since $r(x) = xp(x)$:

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

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

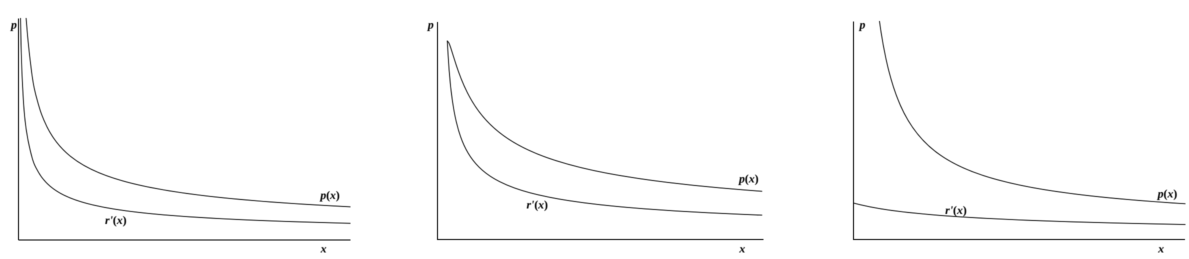
$$p(x) = \frac{\beta}{x}(x-\gamma)^{\frac{\sigma-1}{\sigma}}, \quad 1 < \sigma < \infty,\ x > \gamma\sigma,\ \beta > 0 \tag{2}$$

We are not aware of any previous discussion of the family of inverse demand functions in (2), which expresses expenditure as a power function of consumption relative to a benchmark γ. We detail its properties in Appendix D. Its key property, from (1), 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." It includes CES or isoelastic demands as a special case: when γ equals zero, (2) reduces to $p(x) = \beta x^{-\frac{1}{\sigma}}$, and the elasticity of demand is constant, equal to σ. More generally, the 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.

To give some intuition for the result that CREMR demands link GPF productivity and sales, consider the Pareto case. A Pareto distribution of productivities φ implies that the elasticity of the density of the productivity distribution is constant: if $G(\varphi)$ is Pareto, so $G(\varphi) = 1 - \left(\frac{\varphi}{\underline{\varphi}}\right)^{-k}$, with density function $g(\varphi) = G'(\varphi)$, then the elasticity of density is $\frac{\varphi g'(\varphi)}{g(\varphi)} = -(k+1)$. Similarly, a Pareto distribution of sales, $r = px$, implies that the elasticity

[6] Our approach does not require that the marginal costs be exogenous. They could be chosen endogenously by firms either by optimizing subject to a variable cost function, as in Zhelobodko, Kokovin, Parenti, and Thisse (2012), or as the outcome of investment in R&D.

of the density of the sales distribution is constant: if $F(r) = 1 - \left(\frac{r}{\underline{r}}\right)^{-n}$, with density function $f(r) = F'(r)$, then the elasticity of density is $\frac{rf'(r)}{f(r)} = -(n+1)$. These two log-linear relationships are only consistent with each other if demands also imply a log-linear relationship between firm productivity and firm sales. In a Melitz-type model, productivity is the inverse of marginal cost, which equals marginal revenue. Hence Pareto productivities and Pareto sales are only consistent with each other if there is a log-linear relationship between marginal and total revenue, which is the eponymous defining feature of CREMR demands. To see this slightly more formally, suppose that the distribution of productivity is Pareto with shape parameter k. Then for any two levels of productivity, c_1^{-1} and c_2^{-1}, the ratio of their survival functions (one minus their cumulative probabilities) is $\left(\frac{c_2}{c_1}\right)^k$. Since firms are profit-maximizers, this is also the ratio of the survival functions of marginal revenues, $\left[\frac{r'(x_2)}{r'(x_1)}\right]^k$. But if the elasticity of marginal revenue to sales revenue is constant and equal to $\frac{1}{\sigma-1}$, this in turn equals $\left(\frac{r_2}{r_1}\right)^{\frac{k}{\sigma-1}}$. Since this is true for any arbitrary level of sales, it implies that sales are distributed as a Pareto with scale parameter $n = \frac{k}{\sigma-1}$. This result was derived for the case of Pareto productivities and CES demands by Chaney (2008). (See also Helpman, Melitz, and Yeaple (2004).) The formal proof, a corollary of Proposition 2, shows that it generalizes from CES to CREMR, and that GPF productivities and CREMR demands are necessary as well as sufficient for this outcome.



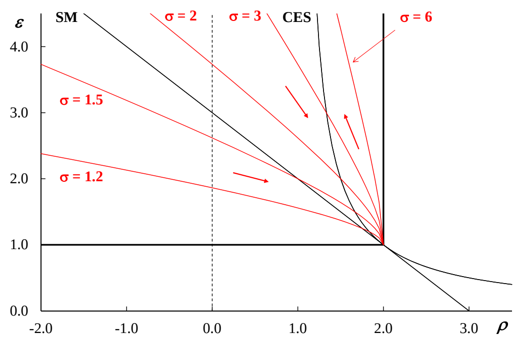
(a) $\gamma = 0$: CES (b) $\gamma > 0$: Subconvex (c) $\gamma < 0$: Superconvex

Figure 2: Examples of CREMR Demand and Marginal Revenue Functions

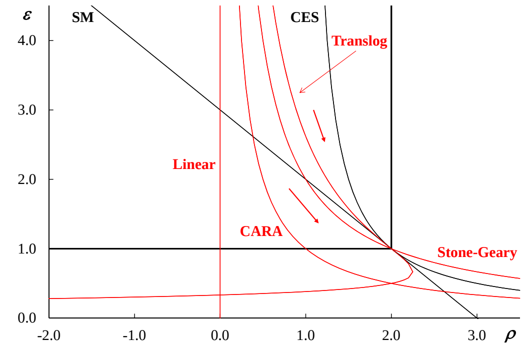
Figure 2 shows three representative inverse demand curves from the CREMR family, along with their corresponding marginal revenue curves. The CES case in panel (a) combines the familiar advantage of analytic tractability with the equally familiar disadvantage of imposing strong and counter-factual properties. In particular, the proportional markup $\frac{p}{c}$ must be the same 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 D 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. In this case the elasticity of demand falls with output, which implies that larger firms have higher markups and that globalization has a pro-competitive effect. 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. Finally, CREMR demands can be rationalized by an additively separable utility function where the sub-utility functions take a hypergeometric form. (For details, see Appendix E.) This is straightforward to simulate, so CREMR demands can also be used as a foundation for quantitative analysis of normative issues.

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 (2013), 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 D that the CREMR demand manifold can be written in closed form as follows:

$$\bar{\rho}(\varepsilon) = 2 - \frac{1}{\sigma - 1} \frac{(\varepsilon - 1)^2}{\varepsilon} \tag{3}$$



(a) CREMR Demands



(b) Some Well-Known Demand Functions

Figure 3: Demand Manifolds for CREMR and Other Demand Functions

Whereas the demand function (2) depends on three parameters, the corresponding demand manifold only depends on σ. Panel (a) of Figure 3 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 Stone-Geary.[7] It is clear that CREMR manifolds, and hence CREMR demand functions, behave very differently from the others. The arrows in Figure 3 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, and their elasticity of demand falls more slowly with convexity as sales rise.

[7] These manifolds are derived in Mrázová and Neary (2013). We confine attention to the admissible region, $\{\varepsilon > 1, \rho < 2\}$, defined as the region 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 (3) 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 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 (2011) for further discussion.

3.2 CREMR and GPF Distributions: Some Special Cases

While the result of the previous sub-section holds for any distributions from the GPF family, it is useful to consider in more detail the Pareto and log-normal cases. Starting with the Pareto, since it is a member of the GPF family of distributions, it follows immediately as a corollary of Proposition 2 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. (In what follows we use $r \sim \mathcal{P}(\underline{r}, n)$ to indicate that r follows a Pareto distribution with threshold parameter $\underline{r}$ and shape parameter n, so $F(r) = 1 - \left(\frac{r}{\underline{r}}\right)^{-n}$.)

Corollary 1. ***Pareto Productivity and Sales Revenue:*** *Any two of the following statements imply the third: 1. Firm productivity $\varphi \sim \mathcal{P}(\underline{\varphi}, k)$; 2. Firm sales revenue $r \sim \mathcal{P}(\underline{r}, n)$; 3. The demand function belongs to the CREMR family in* (2)*; where the parameters are related as follows:*

$$\sigma = \frac{k+n}{n} \quad \Leftrightarrow \quad n = \frac{k}{\sigma - 1} \qquad \textit{and} \qquad \beta = \left(\frac{k+n}{k}\frac{\underline{r}^{\frac{n}{k}}}{\underline{\varphi}}\right)^{\frac{k}{k+n}} \quad \Leftrightarrow \quad \underline{r} = \beta^{\sigma}\left(\frac{\sigma-1}{\sigma}\,\underline{\varphi}\right)^{\sigma-1} \tag{4}$$

Note that the demand parameter γ does not appear in (4), so these expressions hold for all members of the CREMR family, including the CES. This confirms that Corollary 1 extends a result of Chaney (2008), as already noted earlier.

Although it has become customary to assume that actual firm size distributions can be approximated by the Pareto, at least for larger firms, there are other candidate explanations for the pattern of firm sales. Head, Mayer, and Thoenig (2014) and Bee and Schiavo (2014) argue that firm size distribution is better approximated by a log-normal distribution than a Pareto. We have already noted that the log-normal distribution is a special case of the GPF family in Proposition 2. It follows immediately from the proposition that the CREMR relationship $\varphi = \varphi_0 r^E$ is necessary and sufficient for self-reflection in the log-normal case. However, unlike in the Pareto case, this does not imply that all CREMR demand functions are consistent with log-normal productivity and sales. The reason 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.[8] Strictly speaking, this is inconsistent with the log-normal distribution, whose lower bound is zero. We can summarize this result as follows. (We use $r \sim \mathcal{LN}(\mu, s)$ to indicate that r follows a log-normal distribution with location parameter μ and scale parameter s, equal to the mean and standard deviation of the natural logarithm of r. Hence $F(r) = \Phi\left(\frac{\log r - \mu}{s}\right)$, where Φ is the cumulative distribution function of the standard normal distribution.)

Corollary 2. ***Log-Normal Productivity and Sales Revenue:*** *Any two of the following statements imply the third: 1. Firm productivity follows a $\mathcal{LN}(\mu, s)$ distribution; 2. Firm sales follow a $\mathcal{LN}(\mu', s')$ distribution; 3. The demand function is CES: $p(x) = \beta x^{-\frac{1}{\sigma}}$; where the parameters are related as follows:*

$$\sigma = \frac{s+s'}{s} \quad \Leftrightarrow \quad s' = (\sigma - 1)s \qquad \text{and} \qquad \beta = \frac{s+s'}{s'} \exp\left(\frac{s}{s'}\mu' - \mu\right) \quad \Leftrightarrow \quad \mu' = (\sigma - 1)\left[\mu + \log\left(\frac{\beta}{\sigma}\right)\right] \tag{5}$$

Hence, unlike the Pareto case, the only demand function that is exactly compatible with log-normal productivity and sales is the CES. Relaxing the assumption of Pareto productivity in favor of log-normal productivity comes at the expense of ruling out pro-competitive effects. However, in practical applications, where there is a finite interval between the output of the smallest firm and zero, we may not wish to rule out combining log-normal productivity with members of the CREMR family other than the CES.

[8] Since $p'(x) = -\frac{\beta}{\sigma x^2}(x-\gamma)^{-\frac{1}{\sigma}}(x - \gamma\sigma)$, the output of the smallest firm when γ is strictly positive is $\gamma\sigma$, while its sales revenue is $r(x) = \beta\left[\gamma(\sigma-1)\right]^{\frac{\sigma-1}{\sigma}} > 0$. When demands are strictly superconvex, so 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 $r(0) = 0$.

3.3 Self-Reflection of Productivity and Output

The distribution of sales revenue is not the only outcome predicted by models of heterogeneous firms. We can also ask what are the conditions under which output follows the same distribution as productivity. Proposition 2 implies that a necessary and sufficient condition for this form of self-reflection is that the elasticity of productivity with respect to output be constant. This turns out to be related to a different demand family:

$$p(x) = \frac{1}{x}(\alpha + \beta x^{\frac{\sigma-1}{\sigma}}) \tag{6}$$

The demand function in (6) plays the same role with respect to the characteristic of interest, in this case firm output, as the CREMR family does with respect to firm sales. It is necessary and sufficient for a constant elasticity of marginal revenue with respect to output, equal to $\frac{1}{\sigma}$. Hence we call it "CEMR" for "Constant (Output) Elasticity of Marginal Revenue."[9]

Unlike CREMR, there are some precedents for this 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).[10] In particular, the limiting case where σ approaches one is the inverse translog demand function of Christensen, Jorgenson, and Lau (1975). However, except for the CES (the special case when $\alpha = 0$), it bears little resemblance to commonly-used demand functions.[11]

When the common distribution of productivity and output is a Pareto, we can immediately state a further corollary of Proposition 2:

Corollary 3. ***Pareto Productivity and Output:*** *Any two of the following statements imply the third: 1. Firm productivity $\varphi \sim \mathcal{P}(\underline{\varphi}, k)$; 2. Firm output $x \sim \mathcal{P}(\underline{x}, m)$; 3. The de-*

[9] "CEMR" rhymes with "seemer."

[10] For this reason, Mrázová and Neary (2013) called it the "inverse PIGL" class of demand functions.

[11] As shown by Mrázová and Neary (2013), the CEMR demand manifold implies a linear relationship between the convexity and elasticity of demand, passing through the Cobb-Douglas point $(\varepsilon, \rho) = (1, 2)$: $\rho = 2 - \frac{\varepsilon-1}{\sigma}$. For high elasticities (corresponding to small 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 the CREMR manifold becomes $\varepsilon = (2-\rho)(\sigma-1) + 1$.

mand function belongs to the CEMR family (6)*; where the parameters are related as follows:*

$$\sigma=\frac{k}{m}\quad\Leftrightarrow\quad m=\frac{k}{\sigma}\qquad \textit{and}\qquad \beta=\frac{k}{k-m}\,\frac{\underline{x}^{\frac{m}{k}}}{\underline{\varphi}}\quad\Leftrightarrow\quad \underline{x}=\left(\beta\,\frac{\sigma-1}{\sigma}\,\underline{\varphi}\right)^{\sigma}\tag{7}$$

However, when both productivity and output follow a log-normal distribution, we encounter a similar though less extreme restriction on the range of admissible CEMR demand functions to that in the CREMR case of Corollary 2. Now the requirement that output be zero for the smallest firm is only possible if both the parameters α and β in the CEMR demand function (6) are positive. As shown by Mrázová and Neary (2013), this corresponds to the case where CEMR demands are superconvex. By contrast, if either α or β is strictly negative, then demands are strictly subconvex: more plausible in terms of its implications for the distribution of markups, but not compatible with a log-normal distribution of output. Summarizing:

Corollary 4. ***Log-Normal Productivity and Output:*** *Any two of the following statements imply the third: 1. Firm productivity follows a $\mathcal{LN}(\mu,s)$ distribution; 2. Firm output follows a $\mathcal{LN}(\mu',s')$ distribution; 3. The demand function belongs to the superconvex subclass of the CEMR family* (6) *with $\alpha\geq 0$, $\beta\geq 0$, and $\alpha\beta>0$; where the parameters are related as follows:*

$$\sigma=\frac{s'}{s}\quad\Leftrightarrow\quad s'=\sigma s\qquad \textit{and}\qquad \beta=\frac{s'}{s'-s}\exp\left(\frac{s}{s'}\mu'-\mu\right)\quad\Leftrightarrow\quad \mu'=\sigma\left[\mu+\log\left(\beta\frac{\sigma-1}{\sigma}\right)\right]\tag{8}$$

3.4 Self-Reflection of Output and Sales

A final self-reflection corollary of Proposition 2 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.[12] Formally:

[12]Suppose that $x=x_0r(x)^E$. Recalling that $r(x)=xp(x)$, it follows immediately that the demand function must take the CES form.

Corollary 5. ***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 family; 2. The distribution of firm sales revenue r is the same member of the generalized power function family; 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 family class under any demand system other than the CES. Corollary 5 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 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 2 which relates two endogenous firm outcomes rather than an exogenous and an endogenous one.

Taken together, the results of this section show that exactly matching a Pareto or log-normal distribution of firm sales or output, when productivity is assumed to have the same distribution, places strong restrictions on the admissible demand function. The elasticity of marginal revenue with respect to the firm outcome of interest must be constant, and the implied demand function must be consistent with the range of the distribution assumed. However, that leaves open the question of how great an error would be made by using a demand function which does not allow for an exact fit. We address this question in Section 5. First, we turn to consider the implications of different demand functions for the distributions of sales and markups.

4 Inferring Sales and Markup Distributions

The previous section used part (ii) of Proposition 2 to back out the demands implied by assumed distributions of two firm characteristics. In this section we show how part (i) can be used to derive the distributions of firm characteristics given the distribution of productivity and the form of the demand function. Section 4.1 considers the distributions of markups implied by CREMR demands, while Section 4.2 presents the distributions of both sales and markups implied by a number of widely-used demand functions.

4.1 CREMR Markup Distributions

We begin with the case of CREMR demands, since they imply a particularly simple form for the markup distribution. In order to be able to invoke Proposition 2, we need to express productivity as a function of the markup.

The first step is to express output as a function of the markup. In general, with the markup m defined as $\frac{p}{c}$, we can write the markup as a function of output by invoking a standard expression in terms of the elasticity of demand: $m(x) = \frac{\varepsilon(x)}{\varepsilon(x)-1}$. Specializing to the case of CREMR demands, recall from Section 3.1 that the elasticity of demand for CREMR demand functions is $\varepsilon(x) = \frac{x-\gamma}{x-\gamma\sigma}\sigma$. Hence, we can write the CREMR markup as a function of output: $m(x) = \frac{x-\gamma}{x}\frac{\sigma}{\sigma-1}$. We concentrate on the case of subconvex demands (i.e., $\gamma > 0$), which implies that larger firms have higher markups: $m(x) \in \left[\underline{m}, \frac{\sigma}{\sigma-1}\right]$ as $x \in [\underline{x}, \infty]$. Define the relative markup as the markup relative to its maximum value, $\frac{\sigma}{\sigma-1}$, which is the value that obtains under CES preferences with the same value of σ: $\check{m} \equiv \frac{m}{\overline{m}} = \frac{\sigma-1}{\sigma}m \in [\underline{\check{m}}, 1]$. 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: $x(\check{m}) = \frac{\gamma}{1-\check{m}}$.

The next step is to express productivity φ as a function of output. This follows from profit-maximization, which implies that marginal cost φ^{-1} equals marginal revenue, given by equation (26) in Appendix D: $\varphi(x) = \frac{1}{\beta}\frac{\sigma}{\sigma-1}(x-\gamma)^{\frac{1}{\sigma}}$. Finally, combining $\varphi(x)$ and $x(\check{m})$, gives the desired relationship between productivity and the markup:

$$\varphi(\check{m}) = \varphi_0 \left(\frac{\check{m}}{1-\check{m}}\right)^{\frac{1}{\sigma}} \qquad \varphi_0 \equiv \frac{1}{\beta}\,\frac{\sigma}{\sigma-1}\,\gamma^{\frac{1}{\sigma}} \tag{9}$$

Clearly this satisfies Proposition 2's conditions for "Odds Reflection". Hence, if productivity follows any distribution in the GPF class, and if the demand function belongs to the subconvex CREMR family, equation (2) with $\gamma > 0$, then Proposition 2 implies that markups follow the corresponding "GPF-odds" distribution.

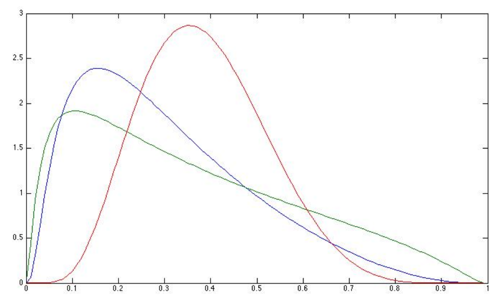


Figure 4: The Log-Normal-Odds Distribution

Once again, we focus on three particularly interesting cases:

1. Pareto: If demands are subconvex CREMR and productivity φ is distributed as a Pareto, so $G(\varphi) = 1 - \underline{\varphi}^k \varphi^{-k}$, then the relative markup must follow a "Pareto-Odds" distribution:

$$F(\check{m}) = 1 - \left(\frac{\underline{\check{m}}}{1-\underline{\check{m}}}\right)^{n'} \left(\frac{\check{m}}{1-\check{m}}\right)^{-n'} \qquad \check{m} \in \{\underline{\check{m}}, 1\} \qquad \check{m} \equiv \frac{m}{\overline{m}},\ \underline{\check{m}} \equiv \frac{\underline{m}}{\overline{m}}. \tag{10}$$

This distribution appears to be new, and may prove useful in future applications. However, it implies that the distribution of markups is U-shaped, which is less in line with the available evidence than the next case we consider.

2. Log-Normal: If demands are subconvex CREMR and productivity follows a log-normal distribution, so $G(\varphi) = \Phi\left[\frac{1}{s}\{\log\varphi - \mu\}\right]$, then the relative markup must follow a

"Log-Normal-Odds" distribution:

$$F(\check{m}) = \Phi\left[\frac{1}{s'}\left\{\log\frac{\check{m}}{1-\check{m}} - \mu'\right\}\right] \tag{11}$$

This distribution has been studied in the statistics literature where it is known as the "Logit-Normal", though we are not aware of a theoretical rationale for its occurrence as here.[13] Figure 4 illustrates some members of this family of distributions. Comparing these with the empirical results illustrated in De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) and Lamorgese, Linarello, and Warzynski (2014), which also exhibit inverted-U-shaped profiles, we can conclude that the log-normal-odds distribution matches the empirical markup distribution extremely well.

3. Fréchet: Finally, if productivity follows a Fréchet distribution and demands are CREMR, then the relative markup must follow a "Fréchet-Odds" distribution. Once again, this distribution appears to be new. It provides an exact characterization of the distribution of profit margins for a firm that sells in many foreign markets, where the distribution of productivity draws across markets follows a Fréchet distribution, as in the model of Tintelnot (2014).

4.2 Other Sales and Markup Distributions

Proposition 2 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, Stone-Geary or linear expenditure system (LES), and translog demands, along with the CREMR results already derived.[14] Combining these with different assumptions about the distribution of productivity, and invoking Proposition

[13]See Johnson (1949) and Mead (1965).

[14]From a firm's perspective, the translog is observationally equivalent to the almost ideal AIDS model of Deaton and Muellbauer (1980).

	$p(x)$ or $x(p)$	$\varphi(r)$ or $\varphi(\check{r})$	$\varphi(m)$ or $\varphi(\check{m})$
CREMR	$\frac{\beta}{x}(x-\gamma)^{\frac{\sigma-1}{\sigma}}$	$\varphi_0 r^{\frac{1}{\sigma-1}}$	$\varphi_0\left(\frac{\check{m}}{1-\check{m}}\right)^{\frac{1}{\sigma}}$
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	$\frac{1}{p}(\gamma-\eta\log p)$	$\varphi_0(r+\eta)\exp\left(\frac{r}{\eta}\right)$	$m\exp\left(m-\frac{\eta+\gamma}{\eta}\right)$

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

2, it is clear that a wide variety of sales and markup distributions are implied.[15] For example, the relationships between productivity 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.[16]

It is clearly desirable to have some way of comparing the distributions implied by these different demand functions with each other and with a given empirical distribution. In the next section we turn to develop tools of this kind.

[15]For parameter restrictions and other details, such as the form of φ_0 (which differs in each case), see Appendix G. Note that in some cases it is desirable to express the results in terms of sales relative to the maximum level, $\check{r} \equiv \frac{r}{\bar{r}}$, just as with CREMR demands the markup distribution is most easily expressed in terms of the relative markup $\check{m}$.

[16]For example, a log-normal distribution of productivity and LES demand imply a log-normal distribution of markups, so providing microfoundations for an assumption made by Epifani and Gancia (2011).

5 Comparing Predicted and Actual Distributions

5.1 From Theory to Calibration

Our approach so far has been to characterize the distributions of firm size and firm markups implied by particular assumptions about the primitives of the model: the structure of demand and the distribution of firm productivities. Results of this kind provide an essential benchmark, but they are not so helpful from a quantitative perspective: they do not tell us by how much this counterfactual distribution departs from a given distribution, whether hypothetical or observed. This is a standard problem encountered in calibration exercises targeting the distribution of firms. Typically, this is done by generating a distribution whose moments match those of the observed distribution. Figure 5 suggests why it may be desirable to follow a different route. It shows the distributions of firms whose productivities are drawn from the same Pareto distribution when the demand they are facing is, respectively, isoelastic and linear. This figure is obtained when the parameters of the demand functions are chosen so that both the mean and the variance of the two distributions are equalized. This suggests that a comparison of the first two moments alone is a poor guide to how "close" two distributions are in practice.[17]

For these reasons, we make use of a different tool to quantify the differences between distributions: the Kullback-Leibler divergence (denoted "KLD" hereafter), introduced by Kullback and Leibler (1951).[18] The next sub-section sketches the information-theoretic background of the KLD, while Section 5.3 shows how it can be applied in our context. Throughout, we concentrate on explaining the distribution of firm sales. Adapting the framework to explain the distribution of output, markups, or any other firm outcome, is straightforward.

[17] There is a separate problem with matching moments for the Pareto distribution. The t'th moment exists if and only if the dispersion parameter k exceeds t; however, empirically, raw data often exhibit values of k that are less than one, so even the mean does not exist.

[18] Other criteria could be used, though none is as satisfactory as the KLD. A first- or second-order stochastic dominance criterion is not informative about the dissimilarity between the two firm size distributions if their cumulative distributions intersect more than once. Similarly, the Kolmogorov-Smirnov test is not very helpful, as it privileges the maximum deviation between the two cumulative distributions, and ignores information about the distributions at other points.

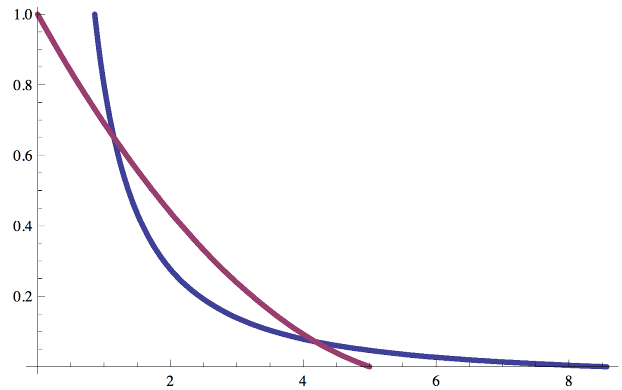


Figure 5: Predicted Size Distributions of Firms from CES and Linear Demands

5.2 Elements of Information Theory

The starting point of information theory is an axiomatic basis for a quantitative measure of the information content of a single draw from a known distribution $F(r)$.[19] It is natural that a measure of information should be additive, non-negative, and inversely related to the probability of the draw. The only function satisfying these requirements is minus the log of the probability: $I(r) = -\log(f(r))$.[20] This in turn leads to the concept of the Shannon entropy of $F(r)$, which is the expected value of information from a single draw:[21]

$$\mathcal{S}_F \equiv E[I(r)] = -\int_{\underline{r}}^{\overline{r}} \log\left(f(r)\right) f(r) dr \tag{12}$$

(See Shannon (1948).) Intuitively, Shannon entropy can be thought of as a measure of the unpredictability or uncertainty about an individual draw implied by the known distribution $F(r)$. In general it ranges from zero to infinity. It equals zero when $F(r)$ is a Dirac distribu-

[19]See Cover and Thomas (2012) for an introduction to information theory. Previous applications of Shannon entropy to economics include the work on inequality by Theil (1967), and the theory of rational inattention developed by Sims (2003), and applied to trade by Dasgupta and Mondria (2014). The KLD has been used as a goodness-of-fit criterion in econometrics (see Vuong (1989), Cameron and Windmeijer (1997) and Ullah (2002)), and empirical demand analysis (see Adams (2013)), but not before to our knowledge in applied theory fields such as international trade.

[20]In information theory it is customary to take all logarithms to base 2, so information is measured in bits. For some theoretical results it is more convenient to use natural logarithms, though most results hold irrespective of the logarithmic base used.

[21]Shannon entropy was first introduced for discrete distributions. The application to continuous distributions is also called "differential entropy".

tion with all its mass concentrated at a single point: in this case, knowing the distribution tells us everything about individual draws, so an extra draw conveys no new information. By contrast, Shannon entropy can be arbitrarily large when $F(r)$ is a uniform distribution:

$$F(r) = \frac{r - \underline{r}}{\overline{r} - \underline{r}}, \ r \in [\underline{r}, \overline{r}] \quad \Rightarrow \quad \mathcal{S}_F = \mathcal{S}_{Uniform} = \log(\overline{r} - \underline{r}) \tag{13}$$

In this case, knowing the distribution conveys no information whatsoever about individual draws, so, as the upper bound $\overline{r}$ becomes arbitrarily large, the same happens to Shannon entropy.

While Shannon entropy measures the expected information gain conveyed by a draw from a single distribution, the KLD measures the information loss when one distribution is used to approximate another one, typically the one observed in the data. Formally, if $F(r)$ is the observed c.d.f. of firms' sales, and $\tilde{F}(r)$ is the distribution used to approximate $F(r)$, then the KLD is defined as follows:

$$\mathcal{D}_{KL}\left(F||\tilde{F}\right) \equiv \int_{\underline{r}}^{\overline{r}} \log\left(\frac{f(r)}{\tilde{f}(r)}\right) f(r)dr \tag{14}$$

To get some intuition for the KLD, it is helpful to rewrite it as follows:

$$\mathcal{D}_{KL}\left(F||\tilde{F}\right) = -\int_{\underline{r}}^{\overline{r}} \log\left(\tilde{f}(r)\right) f(r)dr - \mathcal{S}_F \tag{15}$$

The first term on the right-hand side of (15) measures the cross-entropy between $F(r)$ and $\tilde{F}(r)$. Intuitively, this is a measure of the unpredictability of an individual draw from the benchmark distribution $F(r)$ implied by the tested distribution $\tilde{F}(r)$. Equation (15) thus shows that the KLD equals the difference between the cross-entropy and Shannon entropy. Heuristically, it can be interpreted as the "excess" unpredictability of $F(r)$ implied by $\tilde{F}(r)$ relative to the unpredictability of $F(r)$ implied by itself; or as the informativeness of a draw from $\tilde{F}(r)$ relative to one from $F(r)$. The KLD also has a statistical interpretation:

it equals the expected value of the log likelihood ratio, so choosing the parameters of a distribution to minimize KLD is equivalent to maximizing the likelihood of the sample. By Gibbs' inequality, the KLD is always non-negative, $\mathcal{D}_{KL}\left(F||\tilde{F}\right) \geq 0$, and attains its lower bound of zero if and only if $F(r) = \tilde{F}(r)$ almost everywhere, when the distribution $\tilde{F}(r)$ is completely informative about $F(r)$.

A number of qualifications need to be kept in mind when we use the KLD as a measure of the "closeness" of two distributions. First, the KLD is not symmetric with respect to both distributions: $\mathcal{D}_{KL}\left(\tilde{F}||F\right) \neq \mathcal{D}_{KL}\left(F||\tilde{F}\right)$. Formally, the KLD is a pre-metric, not a metric, and it does not satisfy the triangle inequality. In our application, this does not pose a problem, since it is natural to take the actual firm size distribution as a benchmark, whether it comes from theory or from empirical observation. The role of the KLD is then to quantify how well different candidate methods of calculating a distribution $\tilde{F}(r)$ approximate the "true" distribution $F(r)$: it measures the divergence of $\tilde{F}(r)$ from $F(r)$, not the distance between them.

Second, for the KLD to be well defined, the tested distribution $\tilde{F}(r)$ must have a strictly positive density, $\tilde{f}(r) > 0$, at every point in $[\underline{r}, \overline{r}]$.[22] This can pose problems when we encounter a situation such as that illustrated in Figure 5, where we wish to compare a distribution implied by a demand function (such as the linear) which implies a saturation consumption level with an unbounded distribution such as the Pareto or log-normal. This is not a problem in practical applications, since we can always calibrate demand to fit the upper limit of the observed values of $F(r)$. Even in theoretical contexts, it is an advantage rather than a disadvantage in our context, since it leads us to consider right-truncated distributions. This is a particularly desirable direction to explore in the light of Feenstra (2014), who shows that, without truncation, a Pareto distribution does not allow us to distinguish between the product-variety and pro-competitive gains from trade.

Third, the KLD, like Shannon entropy, attaches the same weight to all observations. In a

[22] The converse is not needed, since by convention $\lim\limits_{f(r)\to 0} f(r)\log\left(f(r)\right) = 0$.

heterogeneous-firms context, we may be more interested in explaining the behavior of large firms, which account for a disproportionate share of total production and exports. One way of implementing this would be to calculate a "weighted KLD", where higher weights are attached to larger firms.[23] A more direct approach, illustrated below, is to see how the KLD behaves as we drop more observations on smaller firms.

5.3 Decomposing the KLD

Because our main focus is on comparing an observed distribution with one predicted by a model, it is helpful to relate the KLD to the elasticities of density of the two distributions. Consider first Shannon entropy. Integrating by parts the definition given in (12) (see Appendix H.1 for details) yields:

$$S_F = -\log f\left(\underline{r}\right) - \int_{\underline{r}}^{\overline{r}} \frac{1 - F\left(r\right)}{r} \frac{rf'\left(r\right)}{f\left(r\right)} dr \tag{16}$$

This shows that Shannon entropy can be decomposed into two terms. The first is the information content of the lower limit of the distribution, i.e., in our application, the information content of the marginal firms. The second equals the integral of the elasticity of the density, $\frac{rf'(r)}{f(r)}$, times the relative survival function, $\frac{1-F(r)}{r}$. The latter is declining in sales, so, when written in this way, Shannon entropy attaches more weight to the elasticities of density of larger firms.[24]

[23] For a discrete version of such a measure, called a "quantitative-qualitative measure of relative information," see Taneja and Tuteja (1984) and Kvålseth (1991). A more satisfactory alternative is the generalization of KLD known as the Rényi divergence of order α, $\alpha \geq 0$ (see Rényi (1959)): $\mathcal{D}_\alpha\big(F||\tilde{F}\big) \equiv \frac{1}{\alpha-1}\log\big(\int_{\underline{r}}^{\overline{r}} \frac{f(r)^\alpha}{\tilde{f}(r)^{\alpha-1}}dr\big)$. The KLD is the limiting case of this as $\alpha \to 1$: $\mathcal{D}_1\big(F||\tilde{F}\big) = \mathcal{D}_{KL}\big(F||\tilde{F}\big)$. For values of α between zero and one, the Rényi divergence weights all possible draws more equally than the KLD, regardless of their probability.

[24] The rate at which the relative survival function declines is one plus the proportional hazard rate: $d\log\left[\frac{1-F(r)}{r}\right] = -\left(1+\frac{rf}{1-F}\right)d\log r$. Equation (38) in Appendix H.1 gives an alternative decomposition of Shannon entropy, where the first term is the information content of the upper limit of the distribution. However, this is less useful in our analytic derivations below, since for many distributions, including the Pareto and the log-normal, $\log f(b) = -\infty$.

Now, we apply the same decomposition to the KLD:

$$\mathcal{D}_{KL}\left(F||\tilde{F}\right) = \log f\left(\underline{r}\right) - \log \tilde{f}\left(\underline{r}\right) + \int_{\underline{r}}^{\overline{r}} \frac{1-F\left(r\right)}{r}\left[\frac{rf'\left(r\right)}{f\left(r\right)} - \frac{r\tilde{f}'\left(r\right)}{\tilde{f}\left(r\right)}\right] dr \tag{17}$$

Once again, this can be decomposed into two terms. The first is the difference between the information contents of the lower limits of the two distributions. The second equals the integral of the difference between their elasticities of density, times the relative survival function, $\frac{1-F(r)}{r}$. Recalling that the latter is declining in sales shows that the KLD attaches less weight to underestimates of the elasticity of density of larger firms.

The decomposition of the KLD in (17) proves particularly insightful when the predicted size distribution is derived from an underlying model of firm behavior. As in Section 3, this comes from a distribution of firm productivity $g(\varphi)$ and a model that links productivity to sales via a function $\varphi(r)$. From the standard result on densities of transformed variables (part (i) of Proposition 1), we can relate the density of the derived distribution of sales to the density of the underlying distribution of firm productivity: $\tilde{f}(r) = g(\varphi(r))\frac{d\varphi}{dr}$. Hence, as we show in Appendix H.2, the elasticity of density of the derived distribution of sales can be expressed in terms of the elasticity of density of the underlying distribution of firm productivity, the revenue elasticity of marginal revenue, $E(r) \equiv \frac{r\varphi'(r)}{\varphi(r)}$, and the elasticity of E with respect to r:

$$\frac{r\tilde{f}'(r)}{\tilde{f}(r)} = \left[\frac{\varphi g'\{\varphi(r)\}}{g\{\varphi(r)\}} + 1\right] E(r) - 1 + \frac{rE'(r)}{E(r)} \tag{18}$$

Substituting this into (17) gives:

$$\begin{aligned}\mathcal{D}_{KL}\left(F||\tilde{F}\right) = & \underbrace{\log f(\underline{r}) - \log\left[g\left(\varphi(\underline{r})\right)\frac{d\varphi}{dr}\bigg|_{\underline{r}}\right]}_{(1)} \\ & + \int_{\underline{r}}^{\overline{r}} \frac{1-F(r)}{r}\left[\underbrace{\left\{\frac{rf'(r)}{f(r)}+1\right\} - \left\{\frac{\varphi g'(\varphi(r))}{g(\varphi(r))}+1\right\}E(r)}_{(2)} - \underbrace{\frac{rE'(r)}{E(r)}}_{(3)}\right] dr\end{aligned} \tag{19}$$

This can be seen as an extension of Proposition 2. That result derived necessary and sufficient conditions for an exact match between the distributions of two firm characteristics when both distributions belonged to the same member of the generalized power function family: the elasticity of one characteristic with respect to the other should be constant, and its value should be consistent with the parameters of the two distributions. Equation (19) goes further and quantifies the information loss when the assumptions of Proposition 2 do not hold. In particular, it identifies three distinct sources of information loss in matching a fitted distribution $\tilde{F}(r)$ to an actual distribution of firm sizes $F(r)$, when the conditions of Proposition 2 do not hold. First is a failure to match the lower end-point of the distribution, $\underline{r}$. Second is the use of an incorrect value of E to link the elasticities of density of the two underlying distributions, $g(\varphi)$ for productivity and $f(r)$ for firm size. And third is a failure to allow for variations in the elasticity E. Each of these three components can be positive or negative, but their sum must be non-negative.

5.4 Quantifying the Information Loss from Incorrect Assumptions

To illustrate the application of equation (19) in the previous section, we show its implications in the benchmark case where both productivity and sales have a Pareto distribution, and demands are of the CREMR type. This eliminates the third source of information loss in (19), since $E' = 0$. However, this does not mean that a perfect calibration is guaranteed, as we shall see.

When $\tilde{F}$ and F are both Pareto with parameters $\tilde{n}$ and n, the KLD can be calculated from equation (17):

$$\mathcal{D}_{KL}(F||\tilde{F}) = \log \frac{n}{\tilde{n}} + \frac{\tilde{n}}{n} - 1 \tag{20}$$

To relate this to the primitive parameters as in (18), recall from Section 3 that with CREMR demands the elasticity of marginal revenue with respect to total revenue, E, equals $\frac{1}{\sigma-1}$, and so, with a Pareto distribution of productivity, the shape parameter for the derived

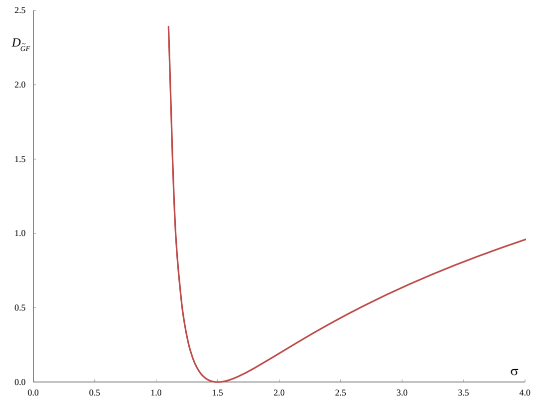
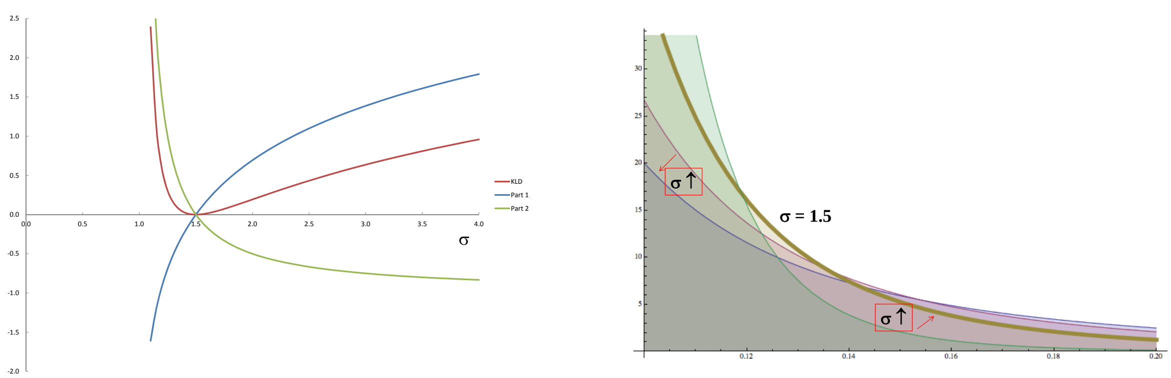


Figure 6: KLD as a Function of σ in the Pareto-CREMR Case

distribution of sales is $\tilde{n} = Ek = \frac{k}{\sigma-1}$. Substituting into (20) gives the KLD decomposition, equation (19), in the Pareto-CREMR case:

$$\mathcal{D}_{KL}(F||\tilde{F}) = \underbrace{\log\frac{n}{k} + \log(\sigma-1)}_{(1)} + \underbrace{\frac{k}{n}\frac{1}{\sigma-1} - 1}_{(2)} \tag{21}$$

This is illustrated in Figure 6 as a function of σ, drawn for values of $k = 1$ and $n = 2$. (The properties of this KLD locus are derived in Appendix H.3.)



(a) Components of KLD

(b) Effects of Changes in σ on Fitted Pareto

Figure 7: Decomposition of KLD in the Pareto-CREMR Case

Figure 6 shows clearly that the information cost of using the "wrong" estimate of σ

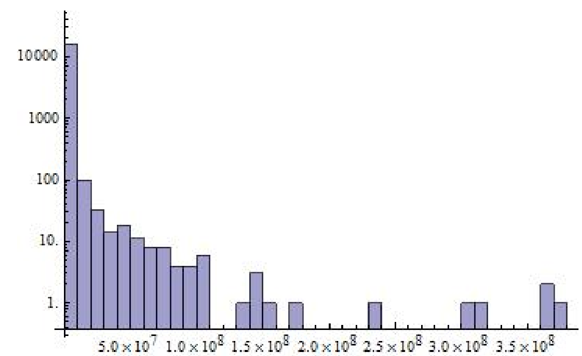
is highly asymmetric. For given values of k and n, the true value of σ equals $\frac{k+n}{n}$. (Recall equation (4).) Given the assumed values of k and n, this equals 1.5, which is the value of σ at which the KLD is minimized. For other values, it is much more sensitive to underestimates than to overestimates of the true value of σ. Why this is so is shown from two different perspectives in Figure 7. Panel (a) shows the two components of the KLD from (21), while panel (b) shows how a higher assumed value of σ affects the location of the predicted distribution relative to that of the true distribution corresponding to $\sigma = 1.5$. Clearly, underestimating σ means overestimating the mass of the smallest firms and underestimating the mass of the larger firms. From (21), the cost of the former is increasing in the log of $\sigma - 1$, whereas the cost of the latter is falling in the reciprocal of $\sigma - 1$. For values of σ below 1.5, the second effect dominates: because the Pareto has an infinite tail, it is more important to fit the larger firms than the smaller ones. This is clear from panel (b), while the numerical values of the components of the KLD in panel (a) show explicitly how the gains and losses in information that come from an increase in σ are traded off against each other.

A further implication of equation (21) is that, with Pareto productivity and CREMR demands, the KLD depends on only one of the three parameters in the CREMR demand function. Figure 6 applies equally well to the CES case (where the CREMR parameter γ is zero) as it does to any other member of the CREMR class. This suggests a further role for the CREMR family in calibrations. To calibrate the size distribution of firms, the only demand parameter that is needed is σ. Hence the values of the other parameters β and γ can be chosen to match other features of the data: γ to match the size distribution of markups across firms, and β to match the level of demand.

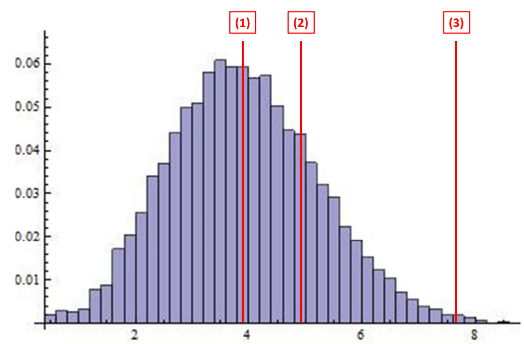
6 Empirical Applications

To illustrate how the KLD can be used to compare the goodness of fit of different assumptions about demand and the distribution of productivity, we end with two empirical applications. The first uses data on French exports to Germany in 2005, drawn from the same source as that used by Head, Mayer, and Thoenig (2014). The second uses firm-level data on Indian sales and markups, as used by De Loecker, Goldberg, Khandelwal, and Pavcnik (2016).

6.1 French Exports to Germany



(a) A First Look: Obviously Pareto?



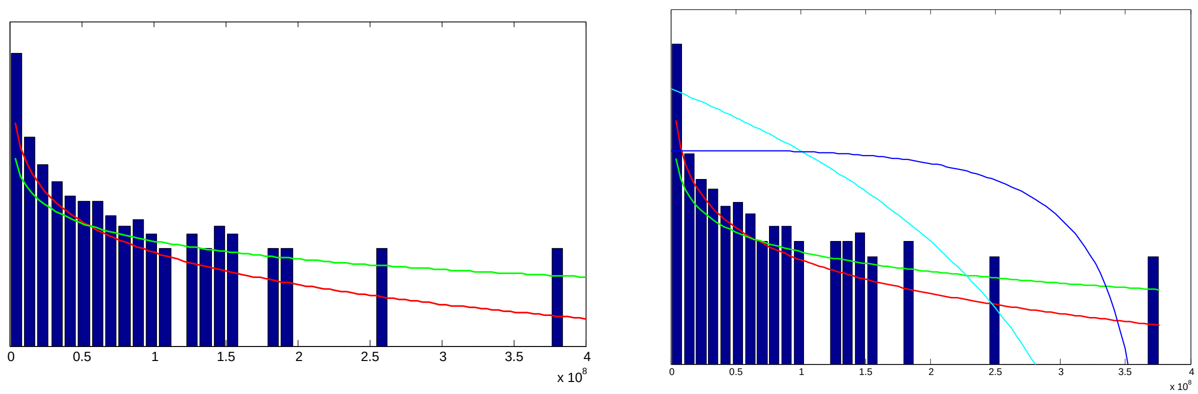
(b) A Second Look: Obviously Log-Normal?

Figure 8: Alternative Perspectives on the Data

The data consists of a 10% representative sample of French exports to Germany in 2005.[25] Figure 8 shows that the data exhibit some typical features of such data sets. When we plot a histogram with the log frequency on the vertical axis and actual sales on the horizontal, as in Panel (a), the long tail is clearly in evidence, and it seems plausible that the data are generated by a Pareto distribution. However, the first bin contains over half the firms, which is brought out more clearly when we plot the actual frequency on the vertical axis and log sales on the horizontal, as in Panel (b). Now the data seem self-evidently log-normal. Yet a

[25] It gives information on 16,119 products exported by 7,928 firms: 2.03 products per firm, all adjusted by an arbitrary constant to preserve confidentiality. We are very grateful to Julien Martin for performing the analysis for us on French Customs data.

third perspective comes from the vertical lines in Panel (b). The line labeled (1) is at median sales, with 50% of firms to the left, but these account for only 0.1% of sales; the line labeled (2) is at 76.7% of firms, but these account for only 1.0% of sales; finally, the line labeled (3) is at 99.6% of firms, which account for only 50% of sales. Thus, we might reasonably conclude that the data are Pareto where it matters, with the top firms dominating.



(a) Pareto and Log-Normal

(b) Linear Demand with Pareto and Log-Normal

Figure 9: KLD-Minimizing Predicted Distributions under Different Assumptions

These subjective considerations provide a poor basis for discriminating between rival views of the best underlying distribution, and justify our turning to use the KLD as a more objective indicator of how well different assumptions fit the data. Figure 9 illustrates some of the candidate distributions. In each case, we choose parameter values for the specification in question that minimize the KLD: recall that this is equivalent to a maximum likelihood estimation of those parameters, conditional on the specification. Panel (a) compares the best-fit Pareto (in green) and log-normal (in red). From Section 3, each of these amounts to assuming that demand is CREMR, and that the underlying productivity distribution is either Pareto or log-normal. (Note that the distribution and demand parameters are not separately identified.) Inspecting the fitted distributions, it is evident that the log-normal matches the smaller firms better, and conversely for the Pareto. Panel (b) of Figure 9 adds the best-fit distributions implied by linear demands in association with either a Pareto (light blue) or log-normal (dark blue) distribution of productivity. These distributions are calculated by combining the relevant productivity distribution with the relationships between productivity and sales implied by linear demands from the middle column of Table 1. (Recall from that table that the linear and LES specifications are observationally equivalent.) Clearly, both imply distributions that are highly log-concave, and that do not fit the data well.

	CREMR/CES	Translog/AIDS	Linear and LES
Pareto	0.0090	0.21	1.40
	$[0.0082, 0.0201]$	$[0.1970, 0.2187]$	$[1.3354, 1.4113]$
Log-Normal	0.0017	2.35	3.68
	$[0.0013, 0.0033]$	$[2.2928, 2.3674]$	$[3.5981, 3.6938]$

Table 2: KLD for Selected Demand Functions and Productivity Distributions
(Bootstrapped 95% Confidence Intervals in Parentheses)

To formalize these comparisons, Table 2 gives the values of the KLD for each of the four cases shown in Figure 9, and also for the translog demand function combined with either Pareto or log-normal productivities. (From a firm's perspective, the translog is observationally equivalent to the almost ideal AIDS model of Deaton and Muellbauer (1980).) Each entry in the table gives the value of the KLD that gives the information loss when the combination of assumptions indicated by the row and column is used to explain the observed distribution of sales. The values in parentheses are bootstrapped 95% confidence intervals. (See Appendix I for details on how these are calculated.) Inspecting these, none of the intervals overlap, implying that the differences between all the KLD values shown are statistically significant at the 5% level.

As we saw in panel (b) of Figure 9, the log-normal matches the smaller firms better, and conversely for the Pareto. With a preponderance of the bins corresponding to smaller firms, it is not surprising that the log-normal does better as measured by the KLD: as shown in the second column of Table 2, it yields a value of 0.0017, considerably lower than the value of 0.0090 for the Pareto. However, the difference between distributions turns out to be much less significant than that between different specifications of demand. The KLD values for the linear/LES specification are much higher than for the CREMR case, as shown in the third column of Table 2, with the Pareto now preferred to the log-normal. A similar pattern is exhibited by the translog/AIDS specification, shown in the final column of Table 2, although now the Pareto does not do so badly.

A different perspective on the values of the KLD statistics in Table 2 is to compare them with the value implied by a uniform distribution of sales. This is in the spirit of 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). With such an uninformative prior, the value of the KLD is 3.5943. Thus we can conclude that linear and translog demands combined with log-normal productivities explain the data about as well as a random explanation, while CREMR and to a lesser extent translog with Pareto do considerably better.

6.2 Indian Sales and Markups

	CREMR	Translog	LES	Linear
A. Sales				
Pareto	0.4110	0.3192	0.6114	0.6114
Log-Normal	0.0205	2.1646	2.7603	2.7603
B. Markups				
Pareto	0.0345	0.0769	0.0749	0.1176
Log-Normal	0.0359	0.0792	0.0619	0.0612

Table 3: KLD for Sales and Markups

The second data set we use gives information on both sales and markups, so for each combination of assumptions on productivity and demand we can calculate the value of the KLD for both sales and markups. (See De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for a detailed description of the data.) The results are given in Table 3 and illustrated

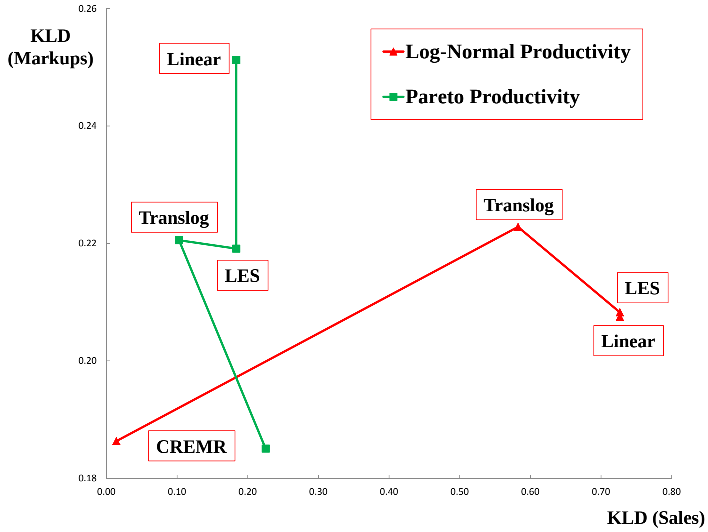


Figure 10: KLD for Sales and Markups

in Figure 10.

With this data set, the results for sales are broadly in line with those from the French data. Assuming a Pareto distribution of productivities, translog demands do somewhat better than Pareto, and linear and LES do not perform much worse. However, the ranking is broadly the same as in the previous case. This is even more pronounced with log-normal productivities: as before, CREMR does best, with translog performing much less well and Linear-LES worst of all.

Of most interest are the results for markups. Here CREMR demands clearly do best, irrespective of the assumed distribution, with translog and LES performing at the same level, and linear doing equally well under Pareto assumptions but less well in the log-normal case. (Recall from Table 1 that linear and LES demands are not separately identified for sales, but they are for markups.) The overwhelming conclusion from these results is that, if we want to fit the distributions of sales and markups in this data set, then the choice between Pareto and log-normal distributions is less important than the choice between CREMR and other demands.

7 Conclusion

This paper has addressed the question of how to explain the distributions of firm size and firm markups using models of heterogeneous firms. We provide a general necessary and sufficient condition for consistency between arbitrary assumptions about the distributions of two firm characteristics and an arbitrary model of firm behavior which relates those two characteristics at the level of an individual firm. 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" family of demands is necessary and sufficient for a Pareto or log-normal distribution of firm productivities to be consistent with a similar distribution of firm sales.

In addition to exact results of this kind, we showed how the Kullback-Leibler divergence can be used to compare a predicted with an observed distribution of firm size. The value of the Kullback-Leibler divergence can be expressed in terms of the difference between the elasticities of density of the two distributions, which in turn can be related to errors in estimating the level and the rate of change of the elasticity of revenue with respect to marginal revenue. Simulations show that the information cost of using the "wrong" parameter to calibrate an observed distribution can be highly asymmetric. Finally, two empirical applications of our approach, to a sample of French exports to Germany and to a dataset of sales and markups for Indian firms, suggest that the choice between Pareto and log-normal distributions is less important than the choice between CREMR and other demands.

While we have concentrated on explaining the distributions of firm sales and markups given assumptions about the distribution of firm productivity, it is clear that our approach has many other potential applications. 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 exact results and our approach to measuring the information cost of incorrect assumptions about behavior should prove useful in many other contexts.

Appendices

A Proof of Proposition 1

To show that (1) and (3) imply (2), let $\tilde{F}(y)$ denote the distribution of y implied by (1) and (3). Since v is strictly increasing from (3), we have $y = v^{-1}(x)$. Therefore the CDF of x is $\tilde{F}[v^{-1}(x)]$. By Assumption 1, it has to coincide with G so:

$$\tilde{F}[v^{-1}(x)] = G(x) \quad \forall x \in (\underline{x}, \bar{x}) \tag{22}$$

Therefore, $\tilde{F}(y) = G[v(y)]$, which is the function assumed in (2), as was to be proved. A similar proof shows that (2) and (3) imply (1).

Next, we wish to prove that (1) and (2) imply (3). We start by picking an arbitrary firm i with characteristics $x(i)$ and $y(i)$. Because $x(i)$ and $y(i)$ are strictly increasing in i, the fraction of firms with characteristics below $x(i)$ and, respectively, $y(i)$, are equal:

$$G[x(i)] = F[y(i)] \quad \forall i \in \Omega \tag{23}$$

Inverting gives $x(i) = G^{-1}[F(y(i))]$. Since this holds for any firm $i \in \Omega$, it follows that $x = v(y) = G^{-1}[F(y)]$, as required.

B Proof of Proposition 2

To show that (1) and (3) imply (2), assume $G(x; \boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2} x^{\theta_2}\right)$, $G_x > 0$, and $x = x_0 h(y)^E$. Then the implied distribution of y is:

$$F(y; \boldsymbol{\theta}) = H\left[\theta_0 + \frac{\theta_1}{\theta_2}\left\{x_0 h(y)^E\right\}^{\theta_2}\right] = H\left[\theta_0 + \frac{\theta_1'}{\theta_2'} h(y)^{\theta_2'}\right] \tag{24}$$

where: $\theta_2' = E\theta_2$ and $\frac{\theta_1'}{\theta_2'} = \frac{\theta_1}{\theta_2}x_0^{\theta_2}$ so $\theta_1' = \frac{\theta_1}{\theta_2}\theta_2' x_0^{\theta_2} = E\theta_1 x_0^{\theta_2}$. Thus (1) and (3) imply (2). A similar proof shows that (2) and (3) imply (1).

Next, to show that (1) and (2) imply (3), assume $G(x;\boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2}x^{\theta_2}\right)$, $G_x > 0$, and $F(y;\boldsymbol{\theta'}) = H\left(\theta_0 + \frac{\theta_1'}{\theta_3'}h(y)^{\theta_2'}\right)$, $F_y > 0$. From part (ii) of Proposition 1, $x = G^{-1}\left[F(y;\boldsymbol{\theta'});\boldsymbol{\theta}\right]$. Inverting $G(x;\boldsymbol{\theta})$ gives $\theta_0 + \frac{\theta_1}{\theta_2}x^{\theta_2} = H^{-1}(G(x;\boldsymbol{\theta}))$, which implies that: $x = \left[\frac{\theta_2}{\theta_1}\left\{H^{-1}(G(x;\boldsymbol{\theta})) - \theta_0\right\}\right]^{\frac{1}{\theta_2}}$. Now substitute $F(y;\boldsymbol{\theta'})$ for $G(x;\boldsymbol{\theta})$:

$$x = \left[\frac{\theta_2}{\theta_1}\left\{H^{-1}\left(H\left(\theta_0 + \frac{\theta_1'}{\theta_3'}h(y)^{\theta_2'}\right)\right) - \theta_0\right\}\right]^{\frac{1}{\theta_2}} = \left[\frac{\theta_2}{\theta_1}\left\{\left(\theta_0 + \frac{\theta_1'}{\theta_2'}h(y)^{\theta_2'}\right) - \theta_0\right\}\right]^{\frac{1}{\theta_2}} = x_0 h(y)^E \tag{25}$$

where: $E = \frac{\theta_2'}{\theta_2}$ and $x_0 = \left(\frac{\theta_2}{\theta_1}\frac{\theta_1'}{\theta_3'}\right)^{\frac{1}{\theta_2}} = \left(\frac{1}{E}\frac{\theta_1'}{\theta_1}\right)^{\frac{1}{\theta_2}}$. Thus (1) and (2) imply (3).

C Generalized Power Function Distributions

Table 4 shows that many well-known distributions are members of the Generalized Power Function family, $G(x;\boldsymbol{\theta}) = H\left(\theta_0 + \frac{\theta_1}{\theta_2}x^{\theta_2}\right)$, introduced in Proposition 2. Hence that proposition 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. (In Section 5.4 we show that the latter condition holds for the truncated Pareto distribution.)

A simple example of a distribution which is not a member of the GPF family is the exponential: $G(x;\boldsymbol{\theta}) = 1 - \exp(-\lambda x)$. This one-parameter distribution does not have the flexibility to match either the sufficiency or the necessity part of Proposition 2. If x is distributed as an exponential and $x = x_0 y^E$, then y is distributed as a Weibull: $F(y;\boldsymbol{\theta'}) = 1 - \exp(-\lambda x_0 y^E)$. Whereas if both x and y are distributed as exponentials, then $x = x_0 y$, i.e., $E = 1$. For similar reasons, the one-parameter version of the Fréchet is not a member of the GPF family, though as Table 4 shows, both its two-parameter version and the three-parameter "Translated Fréchet" (with one of the parameters set equal to θ_0) can be written

	$G(x;\boldsymbol{\theta})$	Support	$H(z)$	θ_0	θ_1	θ_2
Pareto	$1-\left(\frac{x}{\underline{x}}\right)^{-k}$	$[\underline{x},\infty)$	z	1	$k\underline{x}^k$	$-k$
Truncated Pareto	$\frac{1-\underline{x}^k x^{-k}}{1-\underline{x}^k \bar{x}^{-k}}$	$[\underline{x},\bar{x}]$	z	$\frac{1}{1-\underline{x}^k\bar{x}^{-k}}$	$\frac{k\underline{x}^k}{1-\underline{x}^k\bar{x}^{-k}}$	$-k$
Log-normal	$\Phi\left(\frac{\log x-\mu}{s}\right)$	$[0,\infty)$	$\Phi[\log(z)]$	0	$\frac{1}{s}\exp\left(-\frac{\mu}{s}\right)$	$\frac{1}{s}$
Uniform	$\frac{x-\underline{x}}{\bar{x}-\underline{x}}$	$[\underline{x},\bar{x}]$	z	$-\frac{\underline{x}}{\bar{x}-\underline{x}}$	$\frac{1}{\bar{x}-\underline{x}}$	1
Fréchet	$\exp\left[-\left(\frac{x-\mu}{s}\right)^{-\alpha}\right]$	$[\mu,\infty)$	$\exp[-z^{-\alpha}]$	$-\frac{\mu}{s}$	$\frac{1}{s}$	1
Gumbel	$\exp\left[-\exp\left\{-\left(\frac{x-\mu}{s}\right)\right\}\right]$	$(-\infty,\infty)$	$\exp[-\exp\{-z\}]$	$-\frac{\mu}{s}$	$\frac{1}{s}$	1
Reversed Weibull	$\exp\left[-\left(\frac{\mu-x}{s}\right)^{\alpha}\right]$	$(-\infty,\mu]$	$\exp[-z^{\alpha}]$	$\frac{\mu}{s}$	$-\frac{1}{s}$	1

Table 4: Some Members of the Generalized Power Function Family of Distributions

as members of the family.

D 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 (2). To prove sufficiency, note that, from (2), total and marginal revenue are:

$$r(x) \equiv xp(x) = \beta(x-\gamma)^{\frac{\sigma-1}{\sigma}} \qquad r'(x) = p(x) + xp'(x) = \beta\frac{\sigma-1}{\sigma}(x-\gamma)^{-\frac{1}{\sigma}} \tag{26}$$

Combining these gives:

$$r'(x) = \beta^{\frac{\sigma}{\sigma-1}}\frac{\sigma-1}{\sigma}r(x)^{-\frac{1}{\sigma-1}} \tag{27}$$

Hence, the revenue elasticity of marginal revenue is indeed constant, equal to $\frac{1}{\sigma-1}$. For later use it is also useful to express these equations in terms of proportional changes (where a circumflex denotes a logarithmic derivative, so $\hat{r} \equiv \frac{dr}{r}, r > 0$):

$$\left.\begin{array}{l} \hat{r} = \frac{\sigma-1}{\sigma}\frac{x}{x-\gamma}\,\hat{x} \\ \hat{r'} = -\frac{1}{\sigma}\frac{x}{x-\gamma}\,\hat{x} \end{array}\right\} \quad \Rightarrow \quad \hat{r'} = -\frac{1}{\sigma-1}\,\hat{r} \tag{28}$$

To prove necessity, invert equation (1) 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:

$$r(x) = \left[(E+1)\left(\varphi_0^{-1}x - \kappa\right)\right]^{\frac{1}{E+1}} \tag{29}$$

where κ is a constant of integration. Collecting terms, recalling that $r(x) = xp(x)$, gives the CREMR demand system (2), 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 (2), we call the latter CREMR demands.

Next, we wish to derive the demand manifold for CREMR demand functions. Mrázová and Neary (2013) 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. Combining their results leads to an expression for the revenue elasticity of marginal revenue which holds for any demand function:

$$\left.\begin{array}{l} \hat{r} = \frac{\varepsilon-1}{\varepsilon}\,\hat{x} \\ \hat{r'} = -\frac{2-\rho}{\varepsilon-1}\,\hat{x} \end{array}\right\} \quad \Rightarrow \quad \hat{r'} = -\frac{(2-\rho)\varepsilon}{(\varepsilon-1)^2}\,\hat{r} \tag{30}$$

Equating this to (28) leads to the CREMR demand manifold in the text, equation (3). 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.

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 (2013), the expression for the CES locus is: $\rho = \frac{\varepsilon+1}{\varepsilon}$. Eliminating ρ using the CREMR demand manifold (3) and factorizing gives:

$$\rho - \frac{\varepsilon+1}{\varepsilon} = -\frac{(\varepsilon-\sigma)(\varepsilon-1)}{(\sigma-1)\varepsilon} = 0 \tag{31}$$

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 (2013), 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 3.

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 (2013), the expression for the SM locus is: $\rho = 3 - \varepsilon$. Eliminating ρ using the CREMR demand manifold and factorizing gives:

$$\rho + \varepsilon - 3 = \frac{[(\sigma-2)\varepsilon+1](\varepsilon-1)}{(\sigma-1)\varepsilon} = 0 \tag{32}$$

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.

E CREMR Preferences

We seek a specification of preferences which rationalizes CREMR demands. One way of doing this is to assume additively separable preferences, $U = \int_{i\in\Omega} u\{x(i)\}\,\mathrm{d}i$, which implies that $u'\{x(i)\} = \lambda x(i)$, where λ is the marginal utility of income. Integrating the CREMR demand function (2), we can solve for the sub-utility function $u\{x(i)\}$, which takes a hypergeometric form:

$$u\{x(i)\} = \beta\frac{\sigma}{1-\sigma}\,x(i)^{\frac{\sigma-1}{\sigma}}\,{}_2F_1\left(-\frac{\sigma-1}{\sigma}, -\frac{\sigma-1}{\sigma}; \frac{1}{\sigma}; \frac{\gamma}{x(i)}\right) + \kappa \tag{33}$$

Here ${}_2F_1\,(a, b; c; z)$, $|z| < 1$, is the Gaussian hypergeometric function:

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n\,(b)_n}{(c)_n}\frac{z^n}{n!} \tag{34}$$

and $(q)_n$ is the (rising) Pochhammer symbol:

$$(q)_n = \begin{cases} 1 \quad n = 0 \\ q(q+1)...(q+n-1) \quad n > 0 \end{cases} \tag{35}$$

When γ is zero, (33) reduces to the CES utility function, $u\{x(i)\} = \beta\frac{\sigma}{1-\sigma}\,x(i)^{\frac{\sigma-1}{\sigma}} + \kappa$; when γ is positive, so demands are subconvex, utility is positive; and conversely when γ is negative.

Setting κ, the constant of integration in (33), equal to zero implies that $u(0) = 0$. In this case, the utility function always exhibits a taste for diversity. To see this, note that $u(x)$ must be increasing (since otherwise $p(x)$ would not be positive) and concave (since otherwise $p(x)$ would not be decreasing in x). Any concave and differentiable function $u(x)$ is bounded above by its Taylor approximation: $u(x_0) \leq u(x) + (x_0 - x)u'(x)$. Setting $x_0 = 0$ and using the fact that $u(0) = 0$ implies that $0 \leq u(x) - xu'(x)$. Hence the elasticity of utility $\xi(x) \equiv \frac{xu'(x)}{u(x)}$ is always less than one. This in turn implies a taste for diversity in the sense that fixing total consumption $X \equiv nx$, where x is the same for all varieties and n is the measure of varieties, implies that $U = nu(x) = nu\left(\frac{X}{n}\right)$. Logarithmically differentiating with respect to n yields: $\hat{U} = (1 - \xi)\hat{n}$.

F Proofs of Corollaries 1, 2, 3, and 4

Corollaries 1 and 2 (Productivity and Sales with Pareto or Log-Normal):

Proposition 2 holds for any distribution in the generalized power function class. The particular solutions for the constant terms in equations (4) and (5) are derived by substituting the parameters of the Pareto and log-normal distributions into the relevant expressions in Proposition 2. Finally, as discussed in the text, all members of the CREMR class with non-zero γ (i.e., non-zero κ) are, strictly speaking, inconsistent with a log-normal distribution, since they imply that the smallest firm has strictly positive sales revenue.

Corollaries 3 and 4 (Productivity and Output with Pareto or Log-Normal):

In these cases, Proposition 2 implies that productivity must be 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 $r(x)$, with solution:

$$r(x) = \varphi_0^{-1} \frac{x^{1-E}}{1-E} + \kappa \tag{36}$$

where κ is once again a constant of integration. This is the CEMR demand system (6), where $\sigma = \frac{1}{E}$ and $\beta = \frac{1}{\varphi_0(1-E)}$. The final step, as in the case of Corollaries 1 and 2, is to solve for the constant terms when the distributions are either Pareto or log-normal.

G Derivations Underlying Table 1

As in Mrázová and Neary (2013), 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 and general equilibrium. For each demand function, we follow a similar approach to that used with CREMR demands in Sections 3.1 and 4.1: 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 either 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 $\varphi(x) = \frac{1}{\alpha - 2\beta x}$. Finally, the elasticity of demand is $\varepsilon(x) = \frac{\alpha - \beta x}{\beta x}$, so the markup as a function of output is $m(x) = \frac{\alpha - \beta x}{\alpha - 2\beta x}$. 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 $x(m) = \frac{\alpha}{\beta}\frac{m-1}{2m-1}$.

LES: $p(x) = \frac{\delta}{x+\gamma}$, $\gamma > 0, \delta > 0$. We use the inverse demand function rather than the more familiar direct one: $x(p) = \frac{\delta}{p} - \gamma$. Note that, in monopolistic competition, the second-order condition requires that γ be positive, so its usual interpretation as (minus) a subsistence level of consumption is not admissible. Sales revenue is $r(x) = \frac{\delta x}{x+\gamma}$, 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: $\varphi(x) = \frac{(x+\gamma)^2}{\gamma\delta}$. Finally, the elasticity of demand is $\varepsilon(x) = \frac{x+\gamma}{x}$, so the markup as a function of output is $m(x) = \frac{x+\gamma}{\gamma}$; inverting gives $x(m) = \gamma(m-1)$.

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 $r(p) = \gamma - \eta \log p$, which when inverted 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, with: $\varphi_0 = \frac{1}{\exp\left(\frac{\gamma}{\eta}\right)}$. Finally, the elasticity of demand is $\varepsilon(p) = \frac{\eta + \gamma - \eta \log p}{\gamma - \eta \log p}$, so the markup as a function of price is $m(p) = \frac{\eta + \gamma - \eta \log p}{\eta}$; inverting gives $p(m) = \exp\left(\frac{\eta + \gamma}{\eta} - m\right)$.

H Proofs of KLD Properties

H.1 Express KLD in terms of Elasticities of Densities

First, rewrite the definition of Shannon entropy in (12) as $\int_{\underline{r}}^{\overline{r}} udv$, where $u \equiv \log f(r)$, so $du = \frac{f'(r)}{f(r)}dr$, and $dv \equiv f(r)\,dr$, so $v = F(r) + C$, where C is an arbitrary constant of integration. Integrate by parts:

$$S_F = -(1+C)\log f(\overline{r}) + C\log f(\underline{r}) + \int_{\underline{r}}^{\overline{r}} \frac{F(r)+C}{r}\frac{rf'(r)}{f(r)}dr \tag{37}$$

Setting C equal to -1 gives equation (16) in the text, expressed in terms of the lower bound of the distribution, while setting it equal to zero gives an alternative decomposition expressed in terms of the upper bound:

$$S_F = -\log f(\overline{r}) + \int_{\underline{r}}^{\overline{r}} \frac{F(r)}{r}\frac{rf'(r)}{f(r)}dr \tag{38}$$

Repeating this process for the KLD gives in a similar fashion two alternative decompositions, equation (17) in the text and the following:

$$\mathcal{D}_{KL}\left(F||\tilde{F}\right) = \log f(\overline{r}) - \log \tilde{f}(\overline{r}) - \int_{\underline{r}}^{\overline{r}} \frac{F(r)}{r}\left[\frac{rf'(r)}{f(r)} - \frac{r\tilde{f}'(r)}{\tilde{f}(r)}\right]dr \tag{39}$$

H.2 The Elasticity of Density for a Derived Sales Distribution

Recall from Proposition 1 (i) that $\tilde{f}(r) = g(\varphi(r))\frac{d\varphi}{dr}$. Totally differentiating this gives an expression in terms of elasticities:

$$\frac{r\tilde{f}'(r)}{\tilde{f}(r)} = \frac{\varphi(r)g'(\varphi(r))}{g(\varphi(r))}\frac{r\varphi'(r)}{\varphi(r)} + \frac{r\varphi''(r)}{\varphi'(r)} \tag{40}$$

We can relate the second term to the elasticity of marginal revenue with respect to total revenue, $E(r) \equiv \frac{r\varphi'(r)}{\varphi(r)}$:

$$\frac{r\varphi''(r)}{\varphi'(r)} = E(r) - 1 + \frac{rE'(r)}{E(r)} \tag{41}$$

(See Lemma 5 in Mrázová and Neary (2013).) Substituting into (40), the density elasticity of the derived sales distribution $\tilde{F}(r)$ can be written in terms of underlying elasticities as follows:

$$\frac{r\tilde{f}'(r)}{\tilde{f}(r)} = \left[\frac{\varphi(r)g'(\varphi(r))}{g(\varphi(r))} + 1\right] E(r) - 1 + \frac{rE'(r)}{E(r)} \tag{42}$$

Substituting into (17) gives equation (19) in the text. When $G(\varphi)$ is Pareto, $G(\varphi) = 1 - \left(\frac{\varphi}{\underline{\varphi}}\right)^{-k}$, and the elasticity of density is $\frac{\varphi g'(\varphi)}{g(\varphi)} = -(1+k)$. Hence (42) simplifies to the following:

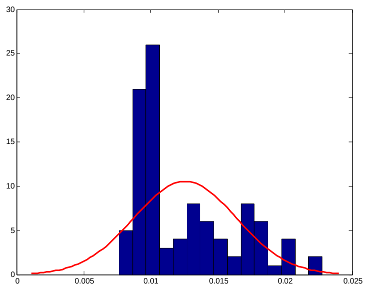
$$\frac{r\tilde{f}'(r)}{\tilde{f}(r)} = -[kE(r) + 1] + \frac{rE'(r)}{E(r)} \tag{43}$$

H.3 The KLD with Pareto and CREMR

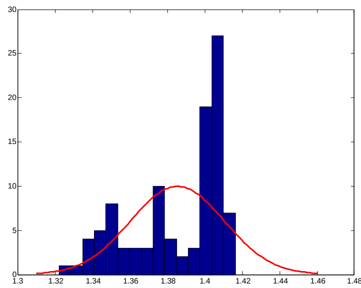
We wish to derive the properties of the KLD in equation (21). The first derivative of this with respect to σ is: $\frac{d\mathcal{D}_{KL}}{d\sigma} = (\sigma-1)^{-1}\left(1 - \frac{\tilde{n}}{n}\right) = (\sigma-1)^{-2}\left(\sigma - \frac{k+n}{n}\right)$. This is positive, and so $\mathcal{D}_{KL}$ is increasing, if and only if $\sigma \geq \frac{k+n}{n}$. The second derivative is: $\frac{d^2\mathcal{D}_{KL}}{d\sigma^2} = -(\sigma-1)^{-2}\left(1 - 2\frac{\tilde{n}}{n}\right) = -(\sigma-1)^{-3}\left(\sigma - \frac{2k+n}{n}\right)$. This is negative, and so $\mathcal{D}_{KL}$ is concave, if and only if $\sigma \geq \frac{2k+n}{n}$.

I Bootstrapped Confidence Intervals for KLD

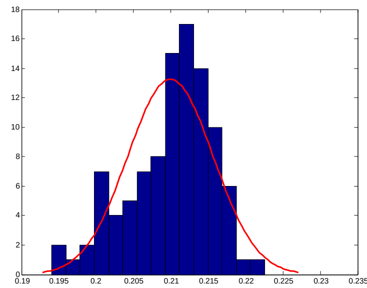
To calculate the confidence intervals given in Table 2, we resampled one hundred times from the observations and, for each sample, calculated the KLD for each of the six combinations of assumptions: Pareto or log-normal distributions of productivity, and CREMR, linear or translog demands. Each of the histograms in Figure 11 shows the frequency distribution of KLD values for one of these combinations, while the smooth curve is a Gaussian kernel



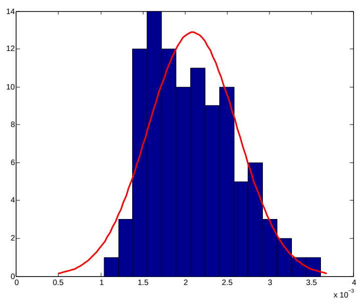
(a) Pareto and CREMR



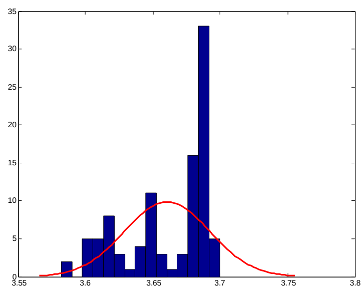
(b) Pareto and Linear



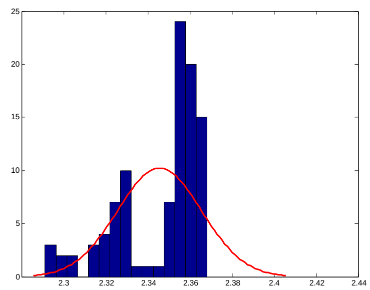
(c) Pareto and Translog



(d) Log-Normal and CREMR



(e) Log-Normal and Linear



(f) Log-Normal and Translog

Figure 11: Frequency and Kernel Distributions of Bootstrapped KLD Values

density fitted to the histogram. Finally, we calculated 95% confidence intervals from the histograms. It is clear from the figures that the sampled KLD values are never symmetric, and in two cases are strongly bimodal, so the calculated confidence intervals should be viewed as approximate only. Nevertheless, they clearly imply that the true underlying distributions do not overlap.

References

Adams, A. (2013): "Nonparametric Demand Systems, Rationality and Welfare Analysis," mimeo, University of Oxford.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2012): "The Elusive Pro-Competitive Effects of Trade," Working paper, Yale University.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012): "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94–130.

Armenter, R., and M. Koren (2014): "A balls-and-bins model of trade," *American Economic Review*, 104(7), 2127–2151.

Bee, M., and S. Schiavo (2014): "Powerless: Gains from Trade when Firm Productivity is not Pareto Distributed," Discussion paper, University of Trento.

Bertoletti, P., and P. Epifani (2014): "Monopolistic Competition: CES Redux?," *Journal of International Economics*, 93(2), 227–238.

Cameron, A. C., and F. A. Windmeijer (1997): "An R-Squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models," *Journal of Econometrics*, 77, 329–342.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4), 1707–1721.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1975): "Transcendental Logarithmic Utility Functions," *American Economic Review*, 65(3), 367–383.

Cover, T. M., and J. A. Thomas (2012): *Elements of Information Theory*. John Wiley and Sons.

Dasgupta, K., and J. Mondria (2014): "Inattentive Importers," Working Paper No. 512, Department of Economics, University of Toronto.

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

Deaton, A., and J. Muellbauer (1980): "An Almost Ideal Demand System," *American Economic Review*, 70(3), 312–326.

Ellison, G., and E. L. Glaeser (1997): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5), 889–927.

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

Feenstra, R. C. (2014): "Restoring the Product Variety and Pro-Competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," NBER Working Paper No. 16796.

Head, K., T. Mayer, and M. Thoenig (2014): "Welfare and Trade Without Pareto," *American Economic Review, Papers and Proceedings*, 104(5), 310–316.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004): "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, 94(1), 300–316.

Johnson, N. (1949): "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36(1-2), 149–176.

Jones, M. (2015): "On Families of Distributions with Shape Parameters," forthcoming in *International Statistical Review*.

Kullback, S., and R. A. Leibler (1951): "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22(1), 79–86.

Kvålseth, T. (1991): "The Relative Useful Information Measure: Some Comments," *Information Sciences*, 56, 35–38.

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

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.

Mead, R. (1965): "A Generalised Logit-Normal Distribution," *Biometrics*, 21(3), 721–732.

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

Melitz, M. J., and G. I. Ottaviano (2008): "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75(1), 295–316.

Mrázová, M., and J. P. Neary (2011): "Selection Effects with Heterogeneous Firms," Discussion Paper No. 588, Department of Economics, University of Oxford.

——— (2013): "Not So Demanding: Preference Structure, Firm Behavior, and Welfare," Discussion Paper No. 691, Department of Economics, University of Oxford.

Muellbauer, J. (1975): "Aggregation, Income Distribution and Consumer Demand," *Review of Economic Studies*, 42(4), 525–543.

Parenti, M., J.-F. Thisse, and P. Ushchev (2014): "Toward a Theory of Monopolistic Ccompetition," CEPR Discussion Paper No. 10014.

Rényi, A. (1959): "On the Dimension and Entropy of Probability Distributions," *Acta Mathematica Academiae Scientiarum Hungarica*, 10(1-2), 193–215.

Shannon, C. E. (1948): "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27(3), 379–423.

Simonovska, I., and M. E. Waugh (2011): "The Elasticity of Trade: Estimates and Evidence," NBER Working Paper No. 16796.

Sims, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

Taneja, H., and R. Tuteja (1984): "Characterization of a Quantitative-Qualitative Measure of Relative Information," *Information Sciences*, 33, 217–222.

Theil, H. (1967): *Economics and Information Theory*. Amsterdam: North-Holland.

Tintelnot, F. (2014): "Global Production with Export Platforms," mimeo., university of chicago.

Ullah, A. (2002): "Uses of Entropy and Divergence Measures for Evaluating Econometric Approximations and Inference," *Journal of Econometrics*, 107(1), 313–326.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012): "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," *Econometrica*, 80(6), 2765–2784.