DISCUSSION PAPER SERIES

DP12044
(v. 2)

SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS

Peter Neary, Monika Mrázová and Mathieu Parenti

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

ISSN 0265-8003

SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS

Peter Neary, Monika Mrázová and Mathieu Parenti

Discussion Paper DP12044
First Published 11 May 2017
This Revision 20 December 2018

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS

Abstract

We derive exact conditions relating the distributions of firm productivity, sales, output, and markups to the form of demand in monopolistic competition. Applications include a new "CREMR" demand function (Constant Revenue Elasticity of Marginal Revenue): it is necessary and sufficient for the distributions of productivity and sales to have the same form (whether Pareto, lognormal, or Fréchet) in the cross section, and for Gibrat's Law to hold over time; it implies a new class of distributions well-suited to capture the dispersion of markups; and it provides a parsimonious fit for the distributions of sales and markups superior to most widely-used alternatives.

JEL Classification: F15, F23, F12

Keywords: CREMR Demands, Gibrat's Law, Heterogeneous Firms, Kullback-Leibler Divergence, Lognormal versus Pareto Distributions, Sales and Markup Distributions

Peter Neary - peter.neary@economics.ox.ac.uk
University of Oxford and CEPR

Monika Mrázová - monika.mrazova@unige.ch
Geneva School of Management and CEPR

Mathieu Parenti - mathieu.parenti@ulb.ac.be
Universite Libre de Bruxelles and CEPR

Acknowledgements
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, to four referees, and to Abi Adams, Costas Arkolakis, Jonathan Dingel, Peter Egger, Basile Grassi, Joe Hirschberg, Oleg Itskhoki, Luca Macedoni, Rosa Matzkin, Marc Melitz, Aviv Nevo, Gianmarco Ottaviano, Steve Redding, Kevin Roberts, Ina Simonovska, Stefan Sperlich, Gonzague Vannoorenberghe, Frank Verboven, Frank Windmeijer, Tianhao Wu, and participants at various conferences and seminars 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.

SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS*

Monika Mrázová†
Geneva and CEPR

J. Peter Neary‡
Oxford, CEPR, and CESifo

Mathieu Parenti§
ULB and CEPR

December 17, 2018

*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, to four referees, and to Abi Adams, Costas Arkolakis, Jonathan Dingel, Peter Egger, Basile Grassi, Joe Hirschberg, Oleg Itskhoki, Luca Macedoni, Rosa Matzkin, Marc Melitz, Aviv Nevo, Gianmarco Ottaviano, Steve Redding, Kevin Roberts, Ina Simonovska, Stefan Sperlich, Gonzague Vannoorenberghe, Frank Verboven, Frank Windmeijer, Tianhao Wu, and participants at various conferences and seminars 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; monika.mrazova@unige.ch.

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

§ECARES, Université Libre de Bruxelles, Av. F.D. Roosevelt, 42, 1050 Brussels, Belgium; 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 monopolistic competition. Applications include a new "CREMR" demand function (Constant Revenue Elasticity of Marginal Revenue): it is necessary and sufficient for the distributions of productivity and sales to have the same form (whether Pareto, lognormal, or Fréchet) in the cross section, and for Gibrat's Law to hold over time; it implies a new class of distributions well-suited to capture the dispersion of markups; and it provides a parsimonious fit for the distributions of sales and markups superior to most widely-used alternatives.

Keywords: CREMR Demands; Gibrat's Law; Heterogeneous Firms; Kullback-Leibler Divergence; Lognormal versus Pareto Distributions; Sales and Markup Distributions.

JEL Classification: F15, F23, F12

1 Introduction

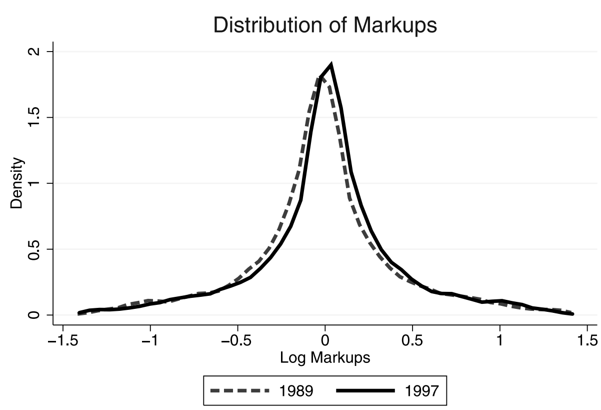
The hypothesis of a representative agent has provided a useful starting point in many fields of economics. However, sooner or later, both intellectual curiosity and the exigencies of matching empirical evidence make it desirable to take account of agent heterogeneity. In many cases, this involves constructing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

Although the pioneering work of Melitz (2003) avoided making specific distributional assumptions, trade models have typically been parameterized in a canonical way, that combines a Pareto distribution of firm productivity on the supply side with CES preferences on the demand side. As shown by Helpman, Melitz, and Yeaple (2004) and Chaney (2008), this combination of assumptions predicts a Pareto distribution of firm sales. This parametrization can be justified on at least two grounds. First is its theoretical tractability, which makes it relatively easy to extend the model to incorporate various real-world features of the global

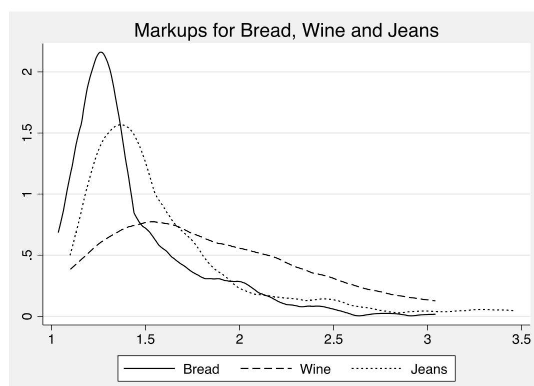
[1] For examples, see Stiglitz (1969), Mirrlees (1971), Krusell and Smith (1998), and Behrens, Duranton, and Robert-Nicoud (2014), respectively.

economy, such as outsourcing, multi-product firms, and global value chains.[2] Second is the empirical regularity that the distribution of firm sales is plausibly close to Pareto, at least in the upper tail. (See, for example, Axtell (2001) and Gabaix (2009).)

However, at least two difficulties arise when this canonical model is confronted with data. First is that not all studies find a Pareto distribution of firm sales, especially if smaller firms are included. Bee and Schiavo (2018) and Head, Mayer, and Thoenig (2014) consider the implications of a lognormal distribution, while Combes, Duranton, Gobillon, Puga, and Roux (2012), and Nigai (2017) explore mixtures and piecewise combinations of Pareto and lognormal, respectively. Analytically, this literature yields a second result: Head, Mayer, and Thoenig (2014) show that lognormal productivities plus CES demands yield a lognormal distribution of firm sales. The parallel between this result and the Helpman-Melitz-Yeaple-Chaney result for the Pareto-CES case is suggestive, but between them they exhaust the extant theoretical literature, which to date gives no guidance on what may happen with other combinations of assumptions.



(a) India, 1989 and 1997



(b) Chile, 2001-2007

Figure 1: Empirical Evidence on Markup Distributions
Sources: De Loecker et al. (2016) and Lamorgese et al. (2014)

A second problem with the canonical framework is that a CES demand function has

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

strong counterfactual implications. In particular, it implies that markups are constant across space and time: in a cross section, all firms should have the same markup in all markets; while, in time series, exogenous shocks such as globalization cannot affect markups and so competition effects will never be observed. Trade economists have been uneasy with these stark predictions for some time, and a number of contributions has explored the implications of relaxing the CES assumption, though to date without considering their implications for sales and markup distributions.[3] However, only recently has it become possible to confront the predictions of CES-based models with data, following the development of techniques for measuring markups that do not impose assumptions about market structure or the functional form of demand. Figure 1(a) from De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) shows that the distribution of markups from a sample of Indian firms is very far from being concentrated at a single value.[4] A possible explanation is that such markup heterogeneity arises from aggregation across sectors with different elasticities of substitution. However, Figure 1(b) from Lamorgese, Linarello, and Warzynski (2014), who use data on Chilean firms, shows that even greater heterogeneity is observed when the data are disaggregated by sector. Taken together, this evidence suggests that markup distributions are far from the Dirac form implied by CES preferences, but to date there is no model of industry equilibrium that would generate such patterns.

In this paper, we provide a general characterization of the problem of explaining the distributions of firm sales 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 particular forms of the distribution of sales revenue, output, or markups. On the other hand, we use the Kullback-Leibler

[3]The implications of preferences other than CES have been considered by Melitz and Ottaviano (2008), Zhelobodko, Kokovin, Parenti, and Thisse (2012), Fabinger and Weyl (2012), Bertoletti and Epifani (2014), Simonovska (2015), Feenstra and Weinstein (2017), Mrázová and Neary (2017), Parenti, Ushchev, and Thisse (2017), Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2018), and Feenstra (2018), among others.

[4]We discuss these data in more detail in Section 7.1 below.

Divergence to quantify the information loss when a predicted distribution fails to match the actual one. We show that applying this criterion 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.

It hardly needs emphasizing that the assumptions made about productivity distributions and demand structure have crucial implications for a wide range of questions. We mention just three. First is the interpretation of the trade elasticity. The elasticity of trade with respect to trade costs is a constant when demands are CES, as shown by Chaney (2008), and this allows a parsimonious expression for the gains from trade in a wide range of canonical models, as shown by Arkolakis, Costinot, and Rodríguez-Clare (2012); see also Melitz and Redding (2015). Similar results hold with non-CES demands if the distribution of firm productivities is Pareto, as shown by Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2018). However, when the distribution of firm productivities is lognormal, the trade elasticity is variable and does not take a simple analytic form even with CES demands, as shown by Head, Mayer, and Thoenig (2014); see also Bas, Mayer, and Thoenig (2017). A second reason why these assumptions matter is the granular origins of aggregate fluctuations. Gabaix (2011) showed that relaxing the continuum assumption implies that the largest firms can have an impact on aggregate fluctuations, and di Giovanni and Levchenko (2012) showed that similar effects arise in open economies. These arguments rely on the assumption that the upper tail of the distribution of firm size is a power law, so understanding the mechanisms that may generate this is a key research question. Thirdly, the interaction of distributional and demand assumptions matters for quantifying the misallocation of resources. The pioneering study of Hsieh and Klenow (2007) showed that close to half the difference in efficiency between China and India on the one hand and the U.S. on the other could be attributed to an inefficient allocation of labor. However, this was under the maintained hypothesis that demands were CES, which, as Dixit and Stiglitz (1977) showed, implies that goods markets are constrained efficient. With non-CES demands, inefficiency may be partly a reflection of goods-market rather than factor-market distortions, with very different implications for welfare-enhancing policies. (See, for example, Epifani and Gancia (2011) and Dhingra and Morrow (2019).) Finally, the links between productivity and demand have been explored in models that consider the distribution not of firm sales but of their growth rates. A central benchmark is Gibrat's Law, or "The Law of Proportionate Effect", which asserts that the rate of growth of a firm is independent of its size. Among the theoretical explanations that have been advanced for the result, it has been shown to hold in models of monopolistic competition with CES preferences by Luttmer (2007, 2011) and Arkolakis (2010a, 2010b, 2016). This raises the question whether Gibrat's Law is consistent with any other demand functions that allow for variable markups.

In all these cases, the assumptions made about the productivity distribution and demand structure matter for key economic issues, yet the existing literature gives little guidance on the implications of relaxing the standard assumptions, nor how best to proceed when the assumptions of the canonical model do not hold. Our paper aims to throw light on both these questions.

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 5 then apply these results to distributions of sales and markups respectively.[5] 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, lognormal 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."[6] The CREMR class has many other

[5] We use "sales" throughout to refer to sales revenue.
[6] "CREMR" rhymes with "dreamer".

desirable properties; in particular, it is necessary and sufficient for Gibrat's Law to hold over time; and it implies a new class of distributions well-suited to capture the dispersion of markups across firms in a flexible but parsimonious way. However, it is very different from most of the non-CES demand systems used in applied economics. We also derive the distributions of markups that 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 6 presents the Kullback-Leibler Divergence (KLD), which has a natural application to evaluating how "close" a predicted distribution comes to an actual one, and shows how this criterion allows us to quantify the cost of using the "wrong" assumptions about demand or technology to calibrate a hypothetical distribution of firm sales or markups. Section 7 illustrates the results of applying the KLD to actual data sets, and shows that it provides a parsimonious fit for the distributions of sales and markups superior to most widely-used alternatives. Finally, Section 8 concludes, while the Appendix contains proofs of propositions as well as further technical details and robustness checks.

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.[7] Formally:

[7] 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. As for the assumption of monotonicity, it is a property of theoretical models only. 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.

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.

The first result restates a standard result in mathematical statistics in 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: $x = v(y)$, $v'(y) > 0$;

where 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* $g(x) = f[v^{-1}(x)]\frac{d[v^{-1}(x)]}{dx}$.

(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.[8] 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

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.

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

characteristics. All that is required is to derive 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, which is given by the following:

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

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

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

The function $H(\cdot)$ is completely general, other than exhibiting the minimal requirements of a probability distribution: $G(x_{min};\boldsymbol{\theta})=0$ and $G(x_{max};\boldsymbol{\theta})=1$, where x_{min} and x_{max} are the bounds of the support of G; and, to be consistent with a strictly positive density function, $G_x>0$, $H(\cdot)$ must satisfy the restriction: $\theta_1 H'>0$. The great convenience of the GPF family given by (1) is that it nests many of the most widely-used distributions in applied economics, including Pareto, truncated Pareto, lognormal, uniform, Fréchet, Gumbel, and Weibull. (See Appendix B for details of members of the GPF family.)

Combining Proposition 1 and Definition 1 gives the following:[9]

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

(1) The distribution of x is a member of the GPF family: $G(x;\boldsymbol{\theta})=H\left(\theta_0+\frac{\theta_1}{\theta_2}x^{\theta_2}\right)$, $G_x>0$;

(2) The distribution of $h(y)$ has the same form as that of x but with different parameters: $F(y;\boldsymbol{\theta'})=G\left[h(y);\boldsymbol{\theta}'\right]=H\left(\theta_0+\frac{\theta_1'}{\theta_2'}h(y)^{\theta_2'}\right)$, $F_y>0$;

[9] The proof is in Appendix C.

(3) x is a power function of $h(y)$: $x = x_0 h(y)^E$;

where 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 $E = \frac{\theta_2'}{\theta_2}$.

Comparing the distributions of x and y in (1) and (2), they are the same member of the GPF family, except that the θ_1 and θ_2 parameters are different, and that y is subject to an arbitrary monotonic transformation $h(y)$. The $h(\cdot)$ function is completely general, and the elements of the parameter vector $\boldsymbol{\theta}$ can take on any values, except in two respects: h must be monotonically increasing from the strict monotonicity 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 distributions.

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.[10] 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 and of the growth of firm sales in Section 4.1. 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 5.

[10] 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).

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 E.

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 2 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.

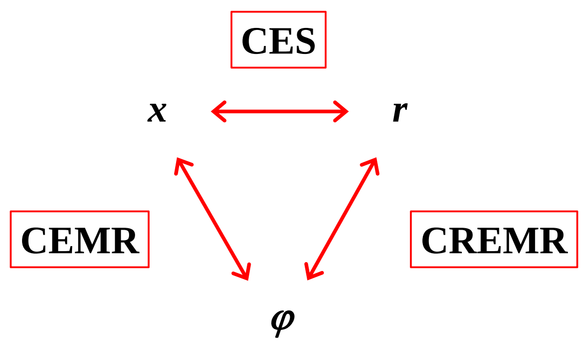


Figure 2: Links Between Firm Characteristics

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 lognormal distribution of productivity coupled with CES demands implies a lognormal 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 lognormal, then they must be related by a power function:

$$\varphi = \varphi_0 r^E \tag{2}$$

To infer the implications of this for demand, we use two properties of a monopolistically competitive equilibrium. First, firms equate marginal cost to marginal revenue, so $\varphi = c^{-1} = \left(\frac{\partial r}{\partial x}\right)^{-1}$.[11] Second, all firms face the same residual demand function, so firm sales conditional on output are independent of productivity φ: $r(x) = xp(x)$ and $\frac{\partial r}{\partial x} = r'(x)$. Combining these with (2) gives a simple differential equation in sales revenue:

$$[r'(x)]^{-1} = \varphi_0 r(x)^E \tag{3}$$

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:

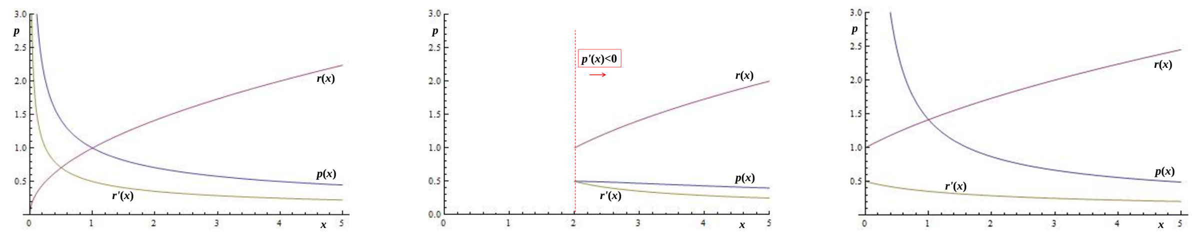
$$p(x) = \frac{\beta}{x}(x-\gamma)^{\frac{\sigma-1}{\sigma}}, \quad 1 < \sigma < \infty,\ x > \gamma\sigma,\ \beta > 0 \tag{4}$$

We are not aware of any previous discussion of the family of inverse demand functions in (4), which expresses expenditure $r(x) = xp(x)$ as a power function of consumption relative to a benchmark γ. We detail its properties in Appendix D. Its key property, from (3), is that the elasticity of marginal revenue with respect to total revenue is constant: $E = \frac{1}{\sigma-1}$. Hence

[11] 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, as in Bustos (2011).

we call it the "CREMR" family, for "Constant Revenue Elasticity of Marginal Revenue." It includes CES demands as a special case: when γ equals zero, (4) 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 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, that it holds for any member of the GPF family not just Pareto, and that GPF productivities and CREMR demands are necessary as well as sufficient.

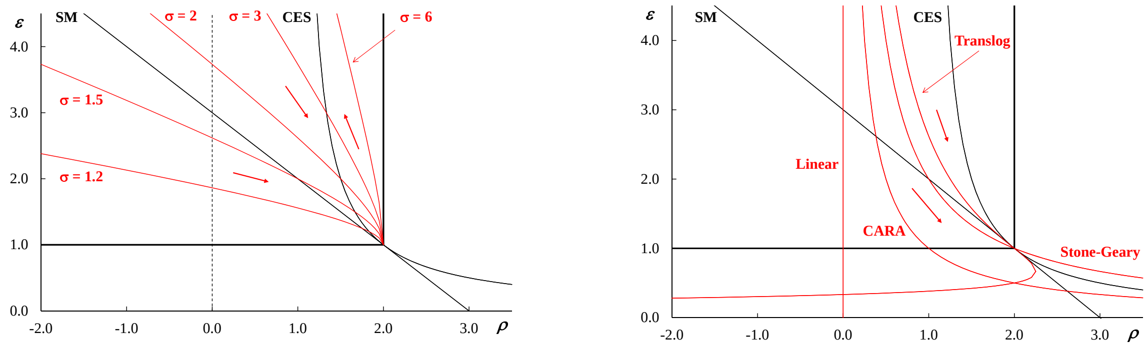


(a) $\gamma = 0$: CES (b) $\gamma > 0$: Subconvex (c) $\gamma < 0$: Superconvex

Figure 3: Examples of CREMR Demand and Marginal Revenue Functions

Figure 3 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, equal to $\frac{\sigma}{\sigma-1}$, for all firms in all markets. By contrast, members of the CREMR family with non-zero values of γ avoid this restriction. Moreover, we show in Appendix 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.

How do CREMR demands compare with other better-known demand systems? Inspecting the demand functions themselves is not so informative, as they depend on three different parameters. Instead, we use the approach of Mrázová and Neary (2017), who show that any



(a) CREMR Demands

(b) Some Well-Known Demand Functions

Figure 4: Demand Manifolds for CREMR and Other Demand Functions

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:

$$\bar{\rho}(\varepsilon) = 2 - \frac{1}{\sigma - 1}\frac{(\varepsilon - 1)^2}{\varepsilon} \tag{5}$$

Whereas the demand function (4) depends on three parameters, the corresponding demand manifold only depends on σ. Panel (a) of Figure 4 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 (or Linear Expenditure System).[12] It is clear that CREMR manifolds, and hence CREMR demand functions, behave very differently from the others. The arrows in Figure 4 denote the direction of movement as sales increase. In the empirically relevant subconvex

[12] These manifolds are derived in Mrázová and Neary (2017). 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 (5) 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 (2019) for further discussion.

region, where demands are less convex than the CES, CREMR demands are more concave at low levels of output (i.e., at high demand elasticities) than any of the others, which, indeed, are approximately linear for small firms. As we move to larger firms, the CREMR elasticity of demand falls more slowly with convexity than any of the others. As for the largest firms, with CREMR demands they asymptote towards a demand function with elasticity equal to σ; whereas with other demand functions the largest firms either hit an upper bound of maximum profitable output (in the linear and CARA cases), or else asymptote to a Cobb-Douglas demand function with elasticity of one (in the translog and Stone-Geary cases).

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 lognormal 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* (4)*; where the parameters are related as follows:*

$$\sigma = \frac{k+n}{n} \quad \Leftrightarrow \quad n = \frac{k}{\sigma - 1} \qquad \textit{and} \qquad \beta = \left(\frac{k+n}{k} \frac{\underline{r}^{\frac{n}{k}}}{\underline{\varphi}}\right)^{\frac{k}{k+n}} \quad \Leftrightarrow \quad \underline{r} = \beta^{\sigma} \left(\frac{\sigma - 1}{\sigma} \underline{\varphi}\right)^{\sigma - 1} \tag{6}$$

Note that the demand parameter γ does not appear in (6), 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 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 (2018) argue that firm size distribution is better approximated by a lognormal distribution than a Pareto. We have already noted that the lognormal 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 lognormal case. However, unlike in the Pareto case, this does not imply that all CREMR demand functions are consistent with lognormal 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.[13] Strictly speaking, this is inconsistent with the lognormal 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 lognormal 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. ***Lognormal 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:*

$$\sigma = \frac{s+s'}{s} \quad \Leftrightarrow \quad s' = (\sigma - 1)s \quad \text{ and } \quad \beta = \frac{s+s'}{s'} \exp\left(\frac{s}{s'}\mu' - \mu\right) \quad \Leftrightarrow \quad \mu' = (\sigma - 1)\left[\mu + \log\left(\frac{\beta}{\sigma}\right)\right] \tag{7}$$

Hence, unlike the Pareto case, the only demand function that is exactly compatible with

[13]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 $r(0) = 0$.

lognormal productivity and sales is the CES. Relaxing the assumption of Pareto productivity in favor of lognormal 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 lognormal productivity with members of the CREMR family other than the CES.

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:

$$p(x) = \frac{1}{x}(\alpha + \beta x^{\frac{\sigma-1}{\sigma}}) \tag{8}$$

The demand function in (8) 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."[14]

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).[15] 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 $\alpha = 0$), CEMR demands bear little resemblance to commonly-used demand functions.[16]

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

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

[16] As shown by Mrázová and Neary (2017), the CEMR demand manifold implies a linear relationship between the convexity and elasticity of demand, passing through the Cobb-Douglas point $(\varepsilon, \rho) = (1, 2)$:

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 demand function belongs to the CEMR family* (8)*; where the parameters are related as follows:*

$$\sigma = \frac{k}{m} \quad \Leftrightarrow \quad m = \frac{k}{\sigma} \qquad \textit{and} \qquad \beta = \frac{k}{k-m} \frac{\underline{x}^{\frac{m}{k}}}{\underline{\varphi}} \quad \Leftrightarrow \quad \underline{x} = \left(\beta \frac{\sigma - 1}{\sigma} \underline{\varphi}\right)^{\sigma} \tag{9}$$

However, when both productivity and output follow a lognormal 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 (8) are positive. As shown by Mrázová and Neary (2017), 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 lognormal distribution of output. Summarizing:

Corollary 4. ***Lognormal 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* (8) *with $\alpha \geq 0$, $\beta \geq 0$, and $\alpha\beta > 0$; where the parameters are related as follows:*

$$\sigma = \frac{s'}{s} \quad \Leftrightarrow \quad s' = \sigma s \qquad \textit{and} \qquad \beta = \frac{s'}{s'-s} \exp\left(\frac{s}{s'}\mu' - \mu\right) \quad \Leftrightarrow \quad \mu' = \sigma \left[\mu + \log\left(\beta \frac{\sigma - 1}{\sigma}\right)\right] \tag{10}$$

$\rho = 2 - \frac{\varepsilon - 1}{\sigma}$. The manifold for the inverse translog special case ($\sigma \to 1$) coincides with the SM locus in Figure 4(b). 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 $\varepsilon = (2 - \rho)(\sigma - 1) + 1$.

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.[17] Formally:

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, output and sales 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.

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

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 6. First, we turn to consider some further features of CREMR demands in Section 4, and then explore their implications for the distribution of markups in Section 5.

4 Applications of CREMR Demands and Preferences

In this section, we explore some further implications of CREMR demands and the preferences that generate them. We show that CREMR demands are necessary and sufficient for Gibrat's Law (i.e., firm growth rates are independent of firm size); we show how CREMR preferences can be used for explicit welfare calculations; and we derive closed-form expressions for the distributions of output in the competitive equilibrium and in the social optimum, showing that, when demand is subconvex, competitive markets encourage too many small firms and not enough large ones relative to the optimum.

4.1 Gibrat's Law

A variety of mechanisms has been proposed to explain the empirical regularity of Gibrat's Law.[18] Early contributions, by Gibrat (1931) himself and by Ijiri and Simon (1974), gave purely stochastic explanations. In particular, if firms are subject to i.i.d. idiosyncratic shocks, these cumulate to give an asymptotic log-normal distribution of firm size, all growing at the same rate. Later work has shown how Gibrat's Law can be derived as an implication of industry equilibrium, when firms are subject to industry-wide as well as idiosyncratic shocks.[19] Much of this work has been carried out under perfectly competitive assumptions, focusing on learning, as in Jovanovic (1982), or differential access to credit, as in Cabral and

[18]For surveys of a large literature, see Sutton (1997) and Luttmer (2010). Gibrat's Law has also been applied to the growth rate of cities. See, for example, Eeckhout (2004). We do not pursue this application here, but it is clear that analogous results to ours can be derived in that case.

[19]As Sutton (1997) points out, different authors have considered shocks to either sales, employment, or assets. In a monopolistically competitive setting, it is natural to assume shocks to productivity, as below.

Mata (2003). The result has also been shown to hold in models of monopolistic competition by Luttmer (2007, 2011) and Arkolakis (2010a, 2010b, 2016). However, these papers assume CES demand. Putting this differently, all models that generate Gibrat's Law to date imply that prices are either equal to or proportional to marginal costs. This raises the question whether Gibrat's Law is consistent with any other demand functions that allow for variable markups.

In fact, we can show that CREMR is the only demand function that allows this:

Proposition 3. *In monopolistic competition, CREMR demands are necessary and sufficient for Gibrat's Law to hold following both idiosyncratic and industry-wide shocks to firm productivity.*

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

Consider first an industry-wide shock. We seek conditions for this to have the same proportionate effect on the sales of all firms. This implies a constant elasticity of sales revenue with respect to marginal cost (recalling that marginal cost is the inverse of productivity). This in turn is equivalent to the CREMR condition for self-reflection that we have already considered, which entails a constant elasticity of marginal revenue with respect to total revenue. Since marginal revenue equals marginal cost, it is immediate that the two conditions are formally identical, though they arise in different contexts: "cross-section" comparisons across firms in the case of self-reflection, by contrast with "time-series" comparisons between the pre- and post-productivity shock equilibria in the case of Gibrat's Law. Hence, invoking Proposition 2, it follows that CREMR demands are necessary and sufficient for Gibrat's Law to hold following industry-wide shocks to firm productivity.

Consider next idiosyncratic shocks to firms' productivity, which can be written as follows:

$$\gamma_{it} - \gamma_{i,t-1} = \epsilon_{it}\gamma_{i,t-1} \tag{11}$$

where ϵ_{it} is i.i.d. Such shocks cumulate to give an asymptotic Lognormal distribution. Equation (11) implies:

$$\log \varphi_{it} = \log \varphi_t + \log \gamma_{0t} + \sum_{t'=0}^{t} \log \epsilon_{it'} \tag{12}$$

As $t \to \infty$, and provided $\log \varphi_t + \log \gamma_{0t}$ is small relative to $\log \varphi_{it}$, the distribution of φ_{it} is approximately Lognormal:

$$\frac{\log \varphi_{it}}{t} \sim N(\mu, \sigma^2) \tag{13}$$

The final step is to recall that productivity equals the inverse of marginal revenue:

$$\varphi_{it} = \varphi_i \gamma_{it} = c_{it}^{-1} = (r'_{it})^{-1} \tag{14}$$

Once again, we can invoke Proposition 2 and conclude that CREMR demands are necessary and sufficient for i.i.d. shocks to productivity to cumulate to give an asymptotic log-normal distribution of sales, with firm growth rates independent of size.

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

4.2 CREMR Preferences

Next, we seek a specification of preferences that 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 $p(i) = \lambda^{-1} u'(x(i))$, where λ is the marginal utility of income. Rewriting the demand shifter in the CREMR demand function (4) as $\beta = \lambda^{-1}\tilde{\beta}$ and integrating yields the sub-utility function $u(x(i))$. This equals the product of two functions, one a CES and the other an augmented hypergeometric, plus a constant of integration κ:

$$u(x(i)) = \tilde{\beta}\frac{\sigma}{\sigma-1}\frac{(x(i)-\gamma)^{\frac{\sigma-1}{\sigma}}}{x(i)}\left(x(i)+\gamma(\sigma-1)\,{}_2F_1\left(1,1,1+\frac{1}{\sigma},\frac{\gamma}{x(i)}\right)\right)+\kappa \tag{15}$$

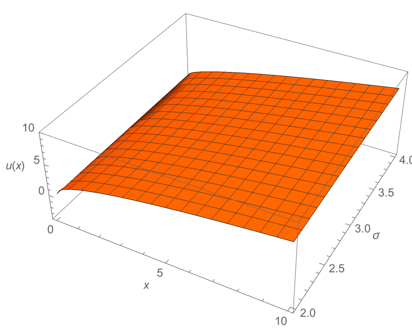
Here ${}_2F_1\,(a,b;c;z)$, $|z|<1$, is the Gaussian hypergeometric function:

$${}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty}\frac{(a)_n\,(b)_n}{(c)_n}\frac{z^n}{n!} \tag{16}$$

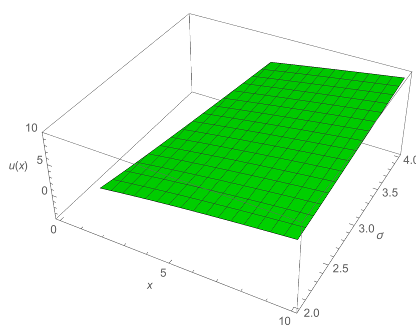
and $(q)_n$ is the (rising) Pochhammer symbol:

$$(q)_n = \frac{\Gamma(q+n)}{\Gamma(q)} \tag{17}$$

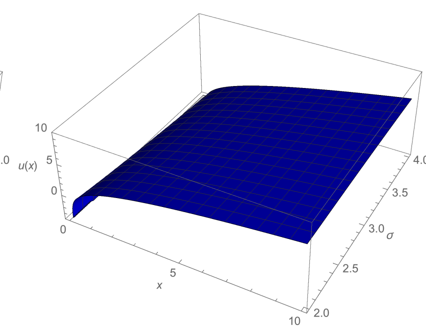
where $\Gamma(q)$ is the gamma function. The final step is to recover the value of λ, which can be done in a standard way.[20]



(a) $\gamma = 0$: CES



(b) $\gamma > 0$: Subconvex



(c) $\gamma < 0$: Superconvex

Figure 5: Examples of CREMR Sub-Utility Functions

When γ is zero, the hypergeometric function equals one, and so (15) reduces to the CES utility function, $u(x(i)) = \beta\frac{\sigma}{\sigma-1}\,x(i)^{\frac{\sigma-1}{\sigma}}+\kappa$. Figure 5 illustrates three sub-utility functions from the CREMR family, each as a function of σ and x, for different values of γ. Panel

[20]Inverting (4) yields the direct demand functions: $x(i) = (u')^{-1}(\lambda p(i))$, which can be combined with the budget constraint to obtain: $\int_{i\in\Omega} p(i)\,(u')^{-1}(\lambda p(i))\,\mathrm{d}i\ = I$ (where I denotes consumer income). Solving this gives λ as a function of prices and income. Note that $x(i)$ cannot be written in closed form, but the marginal utility function is invertible provided the elasticity of demand is positive, i.e., provided $x \in (\max(0,\gamma\sigma),\infty)$.

(a) is the CES case, showing that utility is increasing in σ and increasing and concave in x. The subconcave case in Panel (b) and the superconvex case in Panel (c) (with positive and negative values of γ respectively) deviate from the CES case in ways that correspond to the ways that the corresponding demand functions differ from CES demands in Figure 3.

Setting κ, the constant of integration in (15), 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 $u(x) \geq xu'(x)$. Hence the elasticity of utility $\xi(x) \equiv \frac{xu'(x)}{u(x)}$ is always less than one, which implies that the consumer's preferences exhibit a taste for diversity.[21]

In conclusion, the same techniques can be used to calculate preferences that generate CREMR demands implied by other members of the generalized separability class of Pollak (1972). This is true, for example, of the implicitly additive preferences of Kimball (1990), where the sub-functions corresponding to each good depend on total utility U as well as on the consumption of that good: $\int_{i\in\Omega} \Upsilon(x(i)/U)\, \mathrm{d}i = 1$. A particular advantage of this sub-class is that all members of it exhibit homotheticity. Assuming the demand functions implied by this utility function take a CREMR form, they can be integrated to derive the implied form of $\Upsilon(.)$, which takes a form similar to that in (15). Hence CREMR demands can be rationalized by a number of different forms of generalized separability, where the sub-function corresponding to each good takes a hypergeometric form. Since this is an analytic function, CREMR demands can thus be used as a foundation for quantitative analysis of normative issues.

[21] To see this, assume all varieties have the same price p and so are consumed in equal amounts x, with total expenditure equal to $X = npx$, where n is the measure of varieties. Hence $U = nu(x) = nu\,(X/pn)$. Logarithmically differentiating with respect to n yields: $\hat{U} = (1-\xi)\hat{n}$. Hence, utility is increasing in the number of varieties provided ξ is less than one, as required; i.e., provided an increase in the measure of varieties leads to a utility loss at the intensive margin (equal to ξ) that is less than the gain at the extensive margin (equal to one).

4.3 Quantifying Misallocation at the Microeconomic Level

The results in the previous sub-section make it possible to calculate the aggregate gains from trade under CREMR preferences. A different normative application of CREMR demands is to compare the distributions of outputs across firms in the market equilibrium and in the social optimum. Comparisons between the allocation of resources in a monopolistically competitive market and in the optimum that a social planner would choose have largely focused on the extensive margin, addressing the question of whether the market leads to an under- or over-supply of varieties relative to the social optimum. Dixit and Stiglitz (1977) provided the definitive answer to this question when firms are homogeneous:the market is efficient, in the sense that it supplies the socially optimal number of varieties, and the optimal output of each, if and only if preferences are CES. Feenstra and Kee (2008) showed that the market is also efficient when firms are heterogeneous and the distribution of firm productivities is Pareto, while Dhingra and Morrow (2019) present a general qualitative analysis of the heterogeneous-firm case. Here we focus on a quantitative comparison between the market outcome and the optimal allocation at the intensive margin. In particular, we show how our methods from previous sections can be used to compute the distributions of output in the socially optimal and market outcomes, and we compare the two distributions explicitly in the CREMR case.

We wish to compare the market outcome with the social optimum. Consider first the social planner. Following Dixit and Stiglitz (1977), we assume that the planner cannot make use of lump-sum taxes or subsidies to affect profits. Extending the logic of this assumption to a heterogeneous-firms context, the feasible optimum is a constrained one, where the planner faces the same constraints as the market. In particular, she takes as given the mass of active firms, N_e, the set of goods produced, X, and the productivity threshold, $\bar{\varphi}$, equal to the productivity of a cutoff firm that makes zero profits in the market equilibrium.

The planner maximizes aggregate utility:

$$\int_{i \in X} u(x(i))di = N_e \int_{\bar{\varphi}}^{\infty} u(x(\varphi))g(\varphi)d\varphi \tag{18}$$

subject to the aggregate labor endowment constraint:[22]

$$N_e \int_{\bar{\varphi}}^{\infty} L\varphi^{-1}x(\varphi)g(\varphi)d\varphi + N_e f_e \leq L \tag{19}$$

The first-order condition is:

$$u'(x(\varphi)) = \lambda^* \varphi^{-1} \tag{20}$$

where λ^* is the shadow price of the constraint (19), which we can interpret as the social marginal utility of income; it is defined implicitly by:

$$N_e \int_{\bar{\varphi}}^{\infty} L\varphi^{-1} \left(u'\right)^{-1} (\lambda^* \varphi^{-1}) g(\varphi)d\varphi + N_e f_e = L \tag{21}$$

Hence the planner allocates production across firms according to:

$$\frac{u'(x(\varphi_i))}{u'(x(\varphi_j))} = \frac{\varphi_j}{\varphi_i} \tag{22}$$

which is a standard marginal-cost-pricing rule.

We can say more if the marginal utility of a threshold firm is finite: $u'(\bar{x}) < \infty$, where $\bar{x}$ is the output of a firm with productivity $\bar{\varphi}$. This could be because firms incur fixed costs, or because the demand function implies a positive choke output, as in the case of linear or strictly subconvex CREMR demands. Reexpressing (22) in terms of the output of a typical firm relative to that of a threshold one gives:

$$\frac{u'(x(\varphi))}{u'(\bar{x})} = \frac{\bar{\varphi}}{\varphi} \quad \Rightarrow \quad \varphi^*(x) = \bar{\varphi}\frac{u'(\bar{x})}{u'(x)} \tag{23}$$

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

So the optimal productivity-output relationship depends only on demand.

We are now able to compute the socially optimal distribution of output, using Proposition 2. If the distribution of productivity is given by $G(\varphi)$, then the optimal distribution of output is $F^*(x) = G(\varphi^*(x))$, with $\varphi^*(x)$ determined by (23). This implies a new family of distributions, that we can call the "inverse marginal utility reflection" family:

$$F^*(x) = G\left(\bar{\varphi}\frac{u'(\bar{x})}{u'(x)}\right) \tag{24}$$

Particular members of this family follow if we assume parametric forms for $G(\varphi)$ and $p(x)$. For example, with CREMR demands, $u'(x)$ is proportional to $p(x)$ as given by (4), and $\bar{x} = \gamma\sigma$. (We confine attention to the strictly subconvex case, where $\gamma > 0$.) If, in addition, productivity follows a Pareto distribution we have:

$$F^*(x) = 1 - \left(\frac{(\gamma\sigma - \gamma)^{\frac{\sigma-1}{\sigma}}}{\gamma\sigma}\frac{x}{(x-\gamma)^{\frac{\sigma-1}{\sigma}}}\right)^{-k} \tag{25}$$

If instead productivity follows a Lognormal distribution we have:

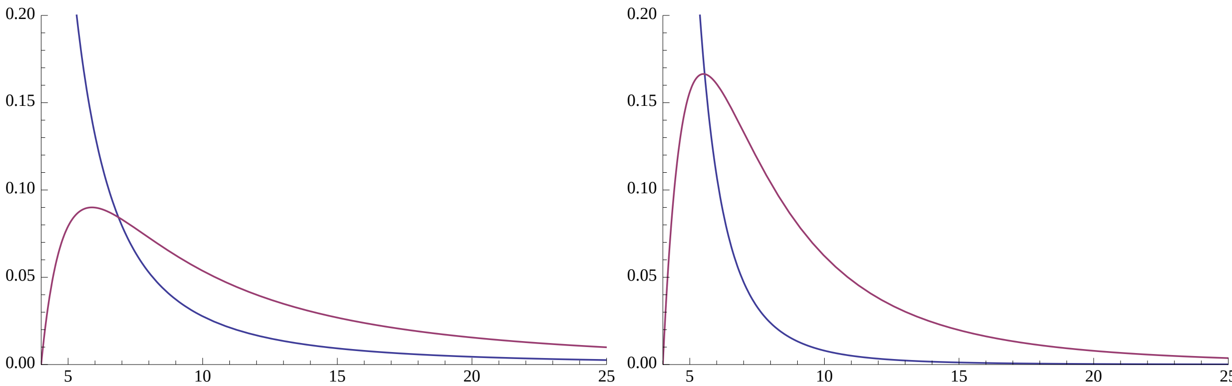
$$F^*(x) = \Phi\left[\frac{1}{s}\left\{A + \log x - \frac{\sigma-1}{\sigma}\log(x-\gamma)\right\}\right] \tag{26}$$

where the constant A is defined as: $A = \log\left(\frac{\bar{\varphi}\beta}{\gamma\sigma}\right) + \frac{\sigma-1}{\sigma}\log(\gamma\sigma - \gamma) - \mu$.

Consider next the market equilibrium. The first-order condition for each firm is that marginal cost should equal marginal revenue, so the productivity-output relationship is:

$$\varphi(x) = \frac{1}{p(x) + xp'(x)} \tag{27}$$

Once again, this implies a new family of distributions, that we can call the "inverse marginal revenue reflection" family. Just as we did in the socially optimal case, we can now compute the distribution of output for any demand function and any distribution of productivities.



(a) Pareto (b) Lognormal

Figure 6: Market versus Socially-Optimal Output Profiles

For example, with strictly subconvex CREMR demands and a Pareto distribution of productivities, the distribution of output in the market equilibrium is:

$$F(x) = 1 - \left(\frac{x - \gamma}{\gamma(\sigma - 1)} \right)^{-\frac{k}{\sigma}} \tag{28}$$

In the Lognormal case, only firms with productivity φ greater than or equal to $\bar{\varphi}$ will produce, so the relevant distribution of productivities is a Lognormal that is left-truncated at $\bar{\varphi}$ where:

$$P(X \leq x \mid X \leq a) = \frac{F(x) - F(a)}{1 - F(a)} \tag{29}$$

So we have:

$$F(x) = \frac{\Phi \left[\frac{1}{s} \left\{ \log \left(\bar{\varphi} \left(\frac{x-\gamma}{\bar{x}-\gamma} \right)^{\frac{1}{\sigma}} \right) - \mu \right\} \right]}{\Phi \left[\frac{1}{s} \left\{ \log \left(\bar{\varphi} \right) - \mu \right\} \right]} - 1 \tag{30}$$

The analytic expressions we have derived allow us to quantify explicitly the pattern of misallocation across firms. Figure 6 illustrates the output profiles in the market equilibrium (in blue) and in the social optimum (in red); panels (a) and (b) give the Pareto and Lognormal cases respectively. The differences between the two panels are unsurprising: both the optimal and the market output profiles are more biased towards smaller firms in the Lognormal case than in the long-tailed Pareto case. More surprising are the similarities between them: in both panels, we see that competitive markets encourage too many small firms and not enough large ones relative to the social optimum.

5 Inferring Sales and Markup Distributions

Section 3 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. In particular, we want to derive the distributions of sales r and markups m, defined as $\frac{p}{c}$. Section 5.1 shows how this is done in general; Section 5.2 considers the distributions of markups implied by CREMR demands; while Section 5.3 presents the distributions of both sales and markups implied by a number of widely-used demand functions.

5.1 Sales and Markup Distributions in General

In order to be able to invoke Proposition 2, we need to express productivity as a function of sales and markups: combining these with the distribution of productivity allows us to derive the implied distributions of sales and markups: $F(r) = G(\varphi(r))$ and $F(m) = G(\varphi(m))$. To see how this works in practice, assume that we know the functional form of the inverse demand function, $p(x)$. (A similar approach is used when we know the direct demand function: see the discussion of the translog case in Appendix G.)

The first step is to express productivity as a function of output. This is straightforward given we know the inverse demand function, since productivity is the inverse of marginal cost, which equals marginal revenue for a profit-maximizing firm. Hence: $\varphi(x) = (r'(x))^{-1} = (p(x) + xp'(x))^{-1}$. Next, to relate output to sales revenue, we need to be able to invert the function $r(x) = xp(x)$. Finally, to express output as a function of the markup, we can use the demand function to calculate the elasticity of demand as a function of output:

$\varepsilon(x) = -p(x)/xp'(x)$, and then write the markup as a function of output by invoking a standard expression in terms of the elasticity of demand: $m(x) = \frac{\varepsilon(x)}{\varepsilon(x)-1}$. This too needs to be inverted to obtain $x(m)$.

5.2 CREMR Markup Distributions

To illustrate how this approach works, we specialize to the case of CREMR demands, which have the added attraction that they imply a particularly simple form for the markup distribution. The first step is to express productivity φ as a function of output, using the expression for CREMR marginal revenue given by equation (42) in Appendix D: $\varphi(x) = \frac{1}{\beta}\frac{\sigma}{\sigma-1}(x-\gamma)^{\frac{1}{\sigma}}$. Next, to relate output to the markup, recall from Section 3.1 that the CREMR elasticity of demand 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: $x(\check{m}) = \frac{\gamma}{1-\check{m}}$. Finally, combining $\varphi(x)$ and $x(\check{m})$, gives the desired relationship between productivity and the markup:

$$\varphi(\check{m}) = \varphi_0 \left(\frac{\check{m}}{1-\check{m}}\right)^{\frac{1}{\sigma}} \qquad \varphi_0 \equiv \frac{1}{\beta}\frac{\sigma}{\sigma-1}\gamma^{\frac{1}{\sigma}} \tag{31}$$

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 (4) with $\gamma > 0$, then Proposition 2 implies that markups follow the corresponding "GPF-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

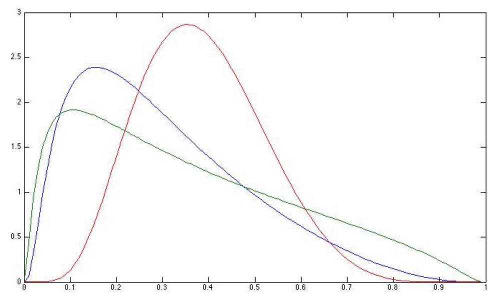


Figure 7: The Lognormal-Odds Distribution

Pareto, so $G(\varphi) = 1 - \underline{\varphi}^k \varphi^{-k}$, then the relative markup must follow a "Pareto-Odds" distribution:

$$F(\check{m}) = 1 - \left(\frac{\underline{\check{m}}}{1-\underline{\check{m}}}\right)^{n'} \left(\frac{\check{m}}{1-\check{m}}\right)^{-n'} \qquad \check{m} \in \{\underline{\check{m}}, 1\} \quad \check{m} \equiv \frac{m}{\overline{m}},\ \underline{\check{m}} \equiv \frac{\underline{m}}{\overline{m}}. \tag{32}$$

where $n' \equiv \frac{k}{\sigma}$ and $\underline{\check{m}} \equiv \frac{\underline{\varphi}^\sigma}{\underline{\varphi}^\sigma + \underline{\varphi}_0^\sigma}$. 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, although the minimum value of the U may lie to the left of the relevant $[0,1]$ interval.

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

$$F(\check{m}) = \Phi\left[\frac{1}{s'}\left\{\log\frac{\check{m}}{1-\check{m}} - \mu'\right\}\right] \tag{33}$$

where: $s' \equiv \sigma s$ and $\mu' \equiv \sigma(\mu - \log\varphi_0)$. 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.[23] Figure 7 illustrates some members of this family of distributions. Comparing these with the empirical results from

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

De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) and Lamorgese, Linarello, and Warzynski (2014) illustrated in Figure 1, which also exhibit inverted-U-shaped profiles, suggests that the lognormal-odds distribution provides a good fit for the empirical markup distribution. Of course, a more precise way of measuring goodness of fit of distributions would be preferable; we will turn to this in the next section.

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 Tintelnot (2017).

5.3 Other Sales and Markup Distributions

	$p(x)$ 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/AI	$\frac{1}{p}(\gamma-\eta\log p)$	$\varphi_0(r+\eta)\exp\left(\frac{r}{\eta}\right)$	$m\exp\left(m-\frac{\eta+\gamma}{\eta}\right)$

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

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.[24] Combining these with different assumptions about the distribution of productivity, and invoking Proposition 2, it is clear that a wide variety of sales and markup distributions are implied.[25] 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.[26]

It is clearly desirable to compare the distributions implied by these different demand functions with each other and with a given empirical distribution. In the remainder of the paper we turn to this task.

6 Comparing Predicted and Actual Distributions

6.1 From Theory to Calibration

So far we have characterized the exact 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 a theoretically-implied distribution departs from a given distribution, whether hypothetical or observed. In the remainder of the paper, we turn to explore the quantitative implications

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

[25]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, $\tilde{r} \equiv \frac{r}{\bar{r}}$, just as with CREMR demands the markup distribution is most easily expressed in terms of the relative markup $\tilde{m}$.

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

of our approach when applied to actual data sets. In particular, we quantify the differences between the actual distributions in the data and a variety of distributions implied by different theoretical models. To measure the "goodness of fit" of different models, we use the Kullback-Leibler divergence (denoted "KLD" hereafter), introduced by Kullback and Leibler (1951). We also present results for the QQ estimator as a robustness check.[27] The next sub-section sketches the theoretical properties of the KLD, while Section 6.3 shows how we operationalize it. To fix ideas, we focus on explaining the distribution of firm sales. Adapting the framework to explain the distribution of output, markups, or any other firm outcome, is straightforward.

6.2 The Kullback-Leibler Divergence

The KLD measures the "information loss" or "relative entropy" when one distribution, F, is used to approximate another, $\tilde{F}$:

$$\mathcal{D}_{KL}(\tilde{F} \mid\mid F(.;\boldsymbol{\theta})) \equiv \int_{\underline{r}}^{\overline{r}} \log\left(\frac{\tilde{f}(r)}{f(r;\boldsymbol{\theta})}\right)\tilde{f}(r)dr \tag{34}$$

In our context, the observed distribution $\tilde{F}(r)$ is the actual distribution of firm sales. As for $F(r;\boldsymbol{\theta})$, it is the theory-consistent distribution of firm sales implied by the assumed underlying distribution of firm productivities, combined with an assumed model of firm behavior.

The KLD has a number of desirable features, the first two of which are well-known. First, it has an axiomatic foundation in information theory: we give further details in Appendix H.1. Second, it has an elegant statistical interpretation: it equals the expected value of the inverse log-likelihood ratio, so choosing the parameter vector $\boldsymbol{\theta}$ to minimize the KLD

[27]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. The Kolmogorov-Smirnov test privileges the maximum deviation between the two cumulative distributions, and ignores information about the distributions at other points. As for matching moments, this does not guarantee a close fit unless many moments are used. Moreover, there is a specific 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.

is asymptotically equivalent to maximizing the likelihood. Third, and new in this paper, is that it links directly with Proposition 2: the KLD in our context can be decomposed to show how it relates to the Revenue Elasticity of Marginal Revenue (REMR) E:

$$\begin{aligned}\mathcal{D}_{KL}(\tilde{F} \,||\, F) = & \underbrace{\log \tilde{f}(\underline{r}) - \log \left[g\left(\varphi(\underline{r})\right) \left.\frac{d\varphi}{dr}\right|_{\underline{r}} \right]}_{(1)} \\ & + \int_{\underline{r}}^{\overline{r}} \frac{1-\tilde{F}(r)}{r} \left[\underbrace{\left\{ \frac{r\tilde{f}'(r)}{\tilde{f}(r)} + 1 \right\} - \left\{ \frac{\varphi g'(\varphi(r))}{g(\varphi(r))} + 1 \right\} E(r)}_{(2)} - \underbrace{\frac{rE'(r)}{E(r)}}_{(3)} \right] dr \end{aligned} \tag{35}$$

Recall that Proposition 2 derived necessary and sufficient conditions for an exact match between the distributions of two firm characteristics when both distributions belong 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 (35) 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 $F(r)$ to an actual distribution of firm sizes $\tilde{F}(r)$, as indicated by the numbered terms in the equation. First is a failure to match the lower end-point of the distribution, $\underline{r}$. Second is a mismatch at each point in the range between the actual elasticity of density of the firm size distribution, $\frac{r\tilde{f}'(r)}{\tilde{f}(r)}$, and that predicted by the assumptions about the productivity distribution and the REMR, $\frac{\varphi g'(\varphi(r))}{g(\varphi(r))}E(r)$. Third is a failure to allow for variations in the REMR, E, itself; i.e., a failure to allow for deviations from part (iii) of Proposition 2. Each of these three components can be positive or negative, but their sum must be non-negative. Appendices H.2 and H.3 give details and applications.

6.3 Operationalizing the KLD

To compare the fit of predicted and actual distributions we use the discrete counterpart of the continuous KLD introduced in Section 6.2. We choose the parameter vector $\boldsymbol{\theta}$ to minimize the KLD between the empirical c.d.f. $\tilde{F}(r)$ defined over the support $[\underline{r}, \bar{r}]$, and the theoretical c.d.f. $F(r;\boldsymbol{\theta})$. Considering the histogram corresponding to $\tilde{F}(r)$, defined over n_b bins with width equal to b, the KLD becomes:

$$\mathcal{D}_{KL}(\tilde{F} \mid\mid F(.;\boldsymbol{\theta})) = \sum_{i=1}^{n_b} \left(\tilde{F}(\underline{r} + i * b) - \tilde{F}(\underline{r} + (i-1) * b)\right) \log\left(\frac{\tilde{F}(\underline{r}+i*b)-\tilde{F}(\underline{r}+(i-1)*b)}{F(\underline{r}+i*b)-F(\underline{r}+(i-1)*b)}\right) \tag{36}$$

We report below $\mathcal{D}_{KL}(\tilde{F} \mid\mid F(.;\hat{\boldsymbol{\theta}}))$, where $\hat{\boldsymbol{\theta}}$ is the parameter vector that minimizes $\mathcal{D}_{KL}(\tilde{F} \mid\mid F(.;\boldsymbol{\theta}))$.

In the next section, we set the number of bins equal to $1,000$. Fortunately, the ranking of different models is not very sensitive to the number of bins considered. As the number of bins increases without bound, our estimator is asymptotically equivalent to the maximum likelihood estimator under the additional constraint that the empirical support of the distribution is included in the one predicted by the theory.[28] As for the units of measurement for the KLD, information scientists typically present values in "bits" (log to base 2) or "nats" (log to base e). Such units have little intuitive appeal in economics. Instead, we present the values of the KLD normalized by the value implied by a uniform distribution of sales.[29] This is an uninformative prior in the spirit of the Laplace principle of insufficient reason; it is analogous to the "dartboard" approach to benchmarking the geographic concentration of manufacturing industry of Ellison and Glaeser (1997), or the "balls and bins" approach to benchmarking the world trade matrix of Armenter and Koren (2014). The value of the KLD is unbounded, but a specification that gave a value greater than that implied by a uniform

[28] When the distribution is lognormal this difference is immaterial as the support consists of $\mathbb{R}+$. This is no longer the case when the distribution is Pareto.

[29] See (62) in Appendix H.1 for the explicit expression.

distribution would be an satisfactory explanation of the data.

7 Fitting the Distributions of Sales and Markups

We turn to show the usefulness of our approach in general, and of CREMR demands in particular, in fitting empirical distributions of both sales and markups. Section 7.1 introduces the firm-level data on Indian sales and markups we use, following De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), and illustrates the results of using the KLD to compare the goodness of fit of different assumptions about demand and the distribution of productivity. Section 7.2 explores how robust are the results to dropping smaller observations. As a further robustness check, Section 7.3 confirms that an alternative criterion for choosing between distributions, the QQ estimator, gives qualitatively similar results to the KLD. Finally, Appendix I gives a second application to French export data: these have the advantage of being more comprehensive but the disadvantage of giving information only on sales and not on markups.

7.1 Indian Sales and Markups

The data set we use has 2,457 firm-product observations on both sales and markups in Indian manufacturing for the year 2001. (See De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for a detailed description of the data, which come from the Prowess data set collected by the Centre for Monitoring the Indian Economy (CMIE).) The sales data are directly observed, while the markup data are estimated, using the so-called "production approach". This approach relies solely on cost-minimization: markups are calculated by computing the gap between the output elasticity with respect to variable inputs and the share of those inputs in total revenue. It is particularly well-suited to our purposes, since it does not impose any restrictions on consumer demand, and it is consistent with a variety of market structures including monopolistic competition. Since the empirical markup distribution has been obtained without making any assumption about functional forms, we can therefore compare the performance of different productivity distributions combined with different demand systems based on the distributions of sales and markups that they imply. The empirical markup distribution was shown in Figure 1(a) above. Observations with negative markups (about 20% of the total) are not included in the sample, as they are inconsistent with steady-state equilibrium behavior by firms. The remaining observations are demeaned by product-year and firm-year fixed effects, so the sample mean equals one by construction.

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

Table 2: KLD for Indian Sales and Markups Compared with Predictions from Selected Demand Functions and Productivity Distributions

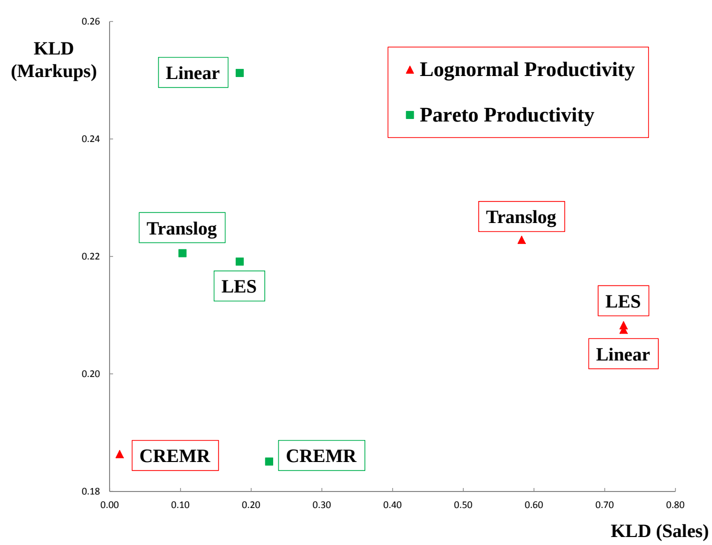


Figure 8: KLD for Indian Sales and Markups

The KLD results are given in Table 2 and illustrated in Figure 8. In each case, we choose parameter values for the specification in question that minimize the KLD: recall that this is asymptotically equivalent to a maximum likelihood estimation of those parameters, conditional on the specification. As already explained, the results are normalized by the KLD for a uniform distribution. Bootstrapping confirms that the differences between them are highly robust, except for that between the predictions of the LES and linear demands for sales: as noted in Section 5.3 these are observationally equivalent. (Appendix J gives details.)

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

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 better under Pareto assumptions but less well in the lognormal case. A clear implication of these results is that the choice between Pareto and lognormal distributions is less important than the choice between CREMR and other demands.

7.2 Robustness to Truncation

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

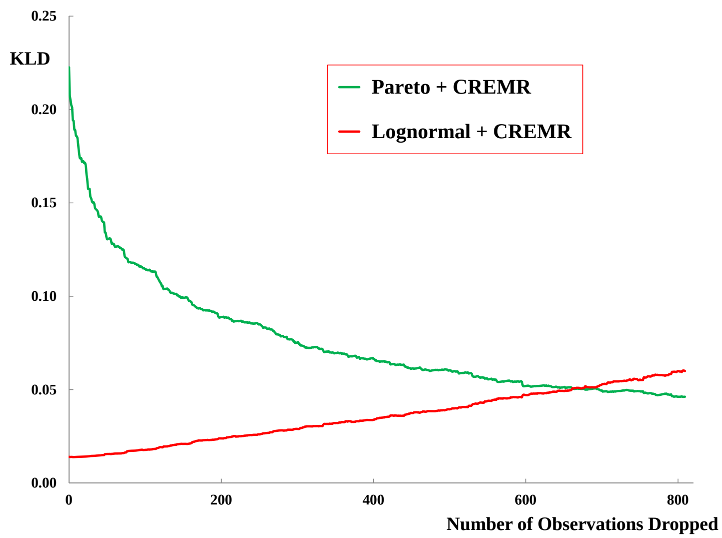


Figure 9: CREMR vs. CREMR: KLD for Indian Sales

Figure 9 compares the KLD for the Pareto and lognormal, conditional on CREMR demands, starting on the left-hand side with all observations (so the values are the same as in Figure 8) and successively dropping up to 809 observations one at a time.[30] Although the curves are not precisely monotonic, the broad picture is clear: conditional on CREMR demands, Pareto does better and lognormal does worse as more and more observations are dropped. The Pareto specification dominates when we drop 663 or more observations: these account for 27% of all firm-product observations, but only 1.2% of total sales.

Figure 10 shows that a similar pattern emerges when we compare the performance of different demand functions in explaining the sales distribution, conditional on a Pareto distribution for productivity. (Note that the horizontal scale differs from that in Figure 9.)

[30] Each KLD value is normalized by the value of the KLD for a uniform distribution corresponding to the number of observations used to calculate it; i.e., excluding the observations dropped. Alternative approaches would make very little difference however, as the KLD value for the uniform varies very little, from 3.9403 with no observations dropped to 3.5598 with 809 observations dropped.

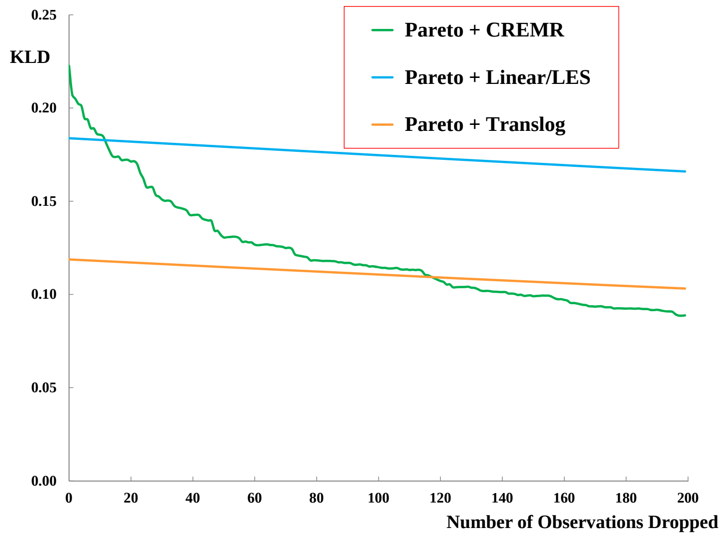


Figure 10: CREMR vs. The Rest, Given Pareto: KLD for Indian Sales

In this case, the CREMR specification overtakes the linear one when we drop 11 or more observations, which account for 0.44% of all firm-product observations, and only 0.0002% of sales. As for the translog, CREMR overtakes it when we drop 118 or more observations, which account for 4.80% of observations, and 0.03% of sales.

These findings confirm that the combination of CREMR demand and Pareto productivities fits the sales data relatively better when the smallest observations are dropped. They also make precise the pattern observed in Figure 15 in Appendix I and in many other datasets, whereby the Pareto assumption outperforms the lognormal in the right tail of the sales distribution. For example, Figure 9 shows that the relevant region in the right tail begins at exactly 663 observations.

7.3 Robustness: The QQ Estimator

A different kind of robustness check is to consider an alternative criterion to the KLD for comparing predicted and actual distributions. Here we consider the QQ estimator, developed by Kratz and Resnick (1996), and previously applied by Head, Mayer, and Thoenig

(2014) and Nigai (2017). Unlike the KLD, this estimator does not have the same desirable theoretical properties as the KLD: in particular, it does not have a maximum likelihood interpretation. However, it is more intuitive, since the QQ distance measure is simply the sum of the squared deviations of the quantiles of the predicted distribution from those of the actual distribution:

$$QQ(\tilde{F} \mid\mid F(\cdot; \boldsymbol{\theta})) = \sum_{i=1}^{n} \left(\log \tilde{q}_i - \log q_i(\boldsymbol{\theta})\right)^2 \tag{37}$$

where $\tilde{q}_i = \tilde{F}^{-1}(i/n)$ is the i'th quantile observed in the data, while $q_i(\boldsymbol{\theta}) = F^{-1}(i/n; \boldsymbol{\theta})$ is the i'th quantile predicted by the theory. The QQ estimator $\hat{\boldsymbol{\theta}}$ is defined as the parameter vector that minimizes the sum of squares $QQ(\tilde{F} \mid\mid F(\cdot; \boldsymbol{\theta}))$ in (37).

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

Table 3: QQ Estimator for Indian Sales and Markups

To implement the QQ estimator we need analytic expressions for the quantiles under each of the eight combinations of assumptions about demand and the distribution of productivity we consider. These are given in Appendix K. We set the number of quantiles n equal to 100. The resulting values of the QQ estimator for Indian sales and markups are given in Table 3, and they are illustrated in Figure 11.

Comparing Table 3 with Table 2, and Figure 11 with Figure 8, it is evident that the results based on the QQ estimator are qualitatively very similar to those for the KLD. In particular, the Pareto assumption gives a better fit for sales than for markups, except in

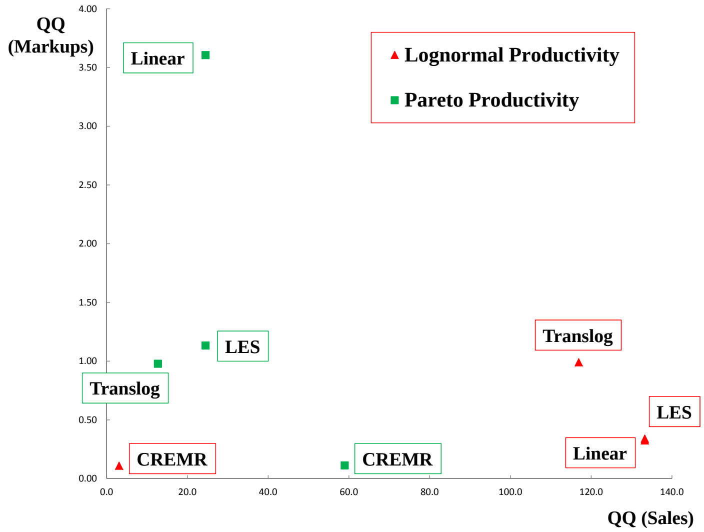


Figure 11: QQ Estimator for Indian Sales and Markups

the CREMR case; while the lognormal assumption tends to give a better fit for markups than for sales. Comparing different demand functions, CREMR demands give a better fit to the markup distribution than any other demands, irrespective of which productivity distribution is assumed. As for sales, the results differ between the Pareto and lognormal cases. Conditional on lognormal, CREMR again performs much better, whereas, conditional on Pareto, it performs least well, with the translog doing best. The only qualitative difference between the results using the two criteria is that with the QQ estimator the translog does somewhat better than the LES in fitting the markup distribution. Overall, we can conclude that the rankings given earlier are not unduly sensitive to our choice of criterion for comparing actual and predicted distributions.

8 Conclusion

This paper has addressed the question of how to explain the distributions of firm size, firm growth rates, 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 lognormal distribution of firm productivities to be consistent with a similar distribution of firm sales. It also allows for variable markups in a parsimonious way, and is necessary and sufficient for Gibrat's Law to hold over time in monopolistic competition.

In addition to exact results of this kind, we have used the Kullback-Leibler divergence to compare the observed distributions of firm sales and firm markups with those predicted by a range of assumptions about the distributions of productivity and the form of demand. Applying our approach to a dataset of sales and markups for Indian firms, suggest that the choice between Pareto and lognormal distributions is less important than the choice between CREMR and other demands. CREMR demands do a particularly good job of explaining markups, and also (conditional on lognormal productivity) give the best fit to sales. Moreover, as in many other data sets, the superior performance of a lognormal relative to a Pareto distribution is sensitive to the presence of very small firms: only a relatively small number of these (accounting for a very small proportion of total sales) needs to be excluded to make the combination of Pareto productivity and CREMR demands the preferred explanation of sales.

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:

$$\tilde{F}[v^{-1}(x)] = G(x) \quad \forall x \in (\underline{x}, \bar{x}) \tag{38}$$

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:

$$G[x(i)] = F[y(i)] \quad \forall i \in \Omega \tag{39}$$

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 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 Definition 1. Hence Proposition 2 can immediately be applied to deduce a constant-elasticity relationship between any two firm characteristics which share any of the distributions in the table, provided the two distributions have compatible supports, and the same value of the parameter θ_0.

	$G(x;\boldsymbol{\theta})$	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	$\frac{1}{1-\underline{x}^k\bar{x}^{-k}}$	$\frac{k\underline{x}^k}{1-\underline{x}^k\bar{x}^{-k}}$	$-k$
Lognormal	$\Phi\left(\frac{\log x-\mu}{s}\right)$	$[0,\infty)$	$\Phi[\log(z)]$	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

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 (used in Eaton and Kortum (2002)) is not a member of the GPF family, though as Table 4 shows, both its two-parameter version (used in many applications of the Eaton-Kortum model) and the three-parameter "Translated Fréchet" (with one of the parameters set equal to θ_0) can be written as members of the family.

C 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:

$$F(y;\boldsymbol{\theta}) = H\left[\theta_0+\frac{\theta_1}{\theta_2}\left\{x_0 h(y)^E\right\}^{\theta_2}\right] = H\left[\theta_0+\frac{\theta_1'}{\theta_2'}h(y)^{\theta_2'}\right] \tag{40}$$

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 $G(x;\boldsymbol{\theta})$:

$$x = \left[\frac{\theta_2}{\theta_1}\left\{H^{-1}\left(H\left(\theta_0 + \frac{\theta_1'}{\theta_3'}h(y)^{\theta_2'}\right)\right) - \theta_0\right\}\right]^{\frac{1}{\theta_2}} = \left[\frac{\theta_2}{\theta_1}\left\{\left(\theta_0 + \frac{\theta_1'}{\theta_2'}h(y)^{\theta_2'}\right) - \theta_0\right\}\right]^{\frac{1}{\theta_2}} = x_0 h(y)^E \tag{41}$$

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).

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 (4). To prove sufficiency, note that, from (4), total and marginal revenue are:

$$r(x) \equiv xp(x) = \beta(x-\gamma)^{\frac{\sigma-1}{\sigma}} \qquad r'(x) = p(x) + xp'(x) = \beta\frac{\sigma-1}{\sigma}(x-\gamma)^{-\frac{1}{\sigma}} \tag{42}$$

Combining these gives:

$$r'(x) = \beta^{\frac{\sigma}{\sigma-1}}\,\frac{\sigma-1}{\sigma}\,r(x)^{-\frac{1}{\sigma-1}} \tag{43}$$

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 $\hat{r} \equiv \frac{dr}{r}, r > 0$):

$$\left.\begin{aligned}\hat{r} &= \tfrac{\sigma-1}{\sigma}\tfrac{x}{x-\gamma}\,\hat{x}\\ \hat{r'} &= -\tfrac{1}{\sigma}\tfrac{x}{x-\gamma}\,\hat{x}\end{aligned}\right\} \;\Rightarrow\; \hat{r'} = -\frac{1}{\sigma-1}\,\hat{r} \tag{44}$$

To prove necessity, invert equation (3) 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:

$$r(x) = \left[(E+1)\left(\varphi_0^{-1} x - \kappa\right)\right]^{\frac{1}{E+1}} \tag{45}$$

where κ is a constant of integration. Collecting terms, recalling that $r(x) = xp(x)$, gives the CREMR demand system (4), 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 (4), we call the latter CREMR demands.

Next, we wish to derive the demand manifold for CREMR demand functions. Mrázová and Neary (2017) 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:

$$\left.\begin{aligned} \hat{r} &= \tfrac{\varepsilon-1}{\varepsilon}\,\hat{x} \\ \hat{r}' &= -\tfrac{2-\rho}{\varepsilon-1}\,\hat{x} \end{aligned}\right\} \quad \Rightarrow \quad \hat{r}' = -\frac{\varepsilon(2-\rho)}{(\varepsilon-1)^2}\,\hat{r} \tag{46}$$

Equating the coefficient of $\hat{r}$ to the corresponding coefficient in the CREMR case, (44), leads to the CREMR demand manifold in the text, equation (5). 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 (2017), the expression for the CES locus is:

$\rho = \frac{\varepsilon+1}{\varepsilon}$. Eliminating ρ using the CREMR demand manifold (5) and factorizing gives:

$$\rho - \frac{\varepsilon+1}{\varepsilon} = -\frac{(\varepsilon-\sigma)(\varepsilon-1)}{(\sigma-1)\varepsilon} = 0 \tag{47}$$

Given $1 < \sigma \leq \infty$, this expression is zero, and so every CREMR manifold intersects the CES locus, at two points. One is at $\{\varepsilon, \rho\} = \{1, 2\}$, implying that all CREMR demand manifolds must pass through the Cobb-Douglas point. The other is at $\{\varepsilon, \rho\} = \{\sigma, 1 + \frac{1}{\sigma}\}$. Hence every CREMR demand manifold lies strictly within the superconvex region (where $\rho > \frac{\varepsilon+1}{\varepsilon}$) for $\sigma > \varepsilon > 1$, and strictly within the subconvex region for $\varepsilon > \sigma$. The condition for superconvexity, $\varepsilon \leq \sigma$, can be reexpressed in terms of γ by using the fact that the elasticity of demand is $\varepsilon = \frac{x-\gamma}{x-\gamma\sigma}\sigma$. Substituting and recalling that σ must be strictly greater than one, we find that CREMR demands are superconvex if and only if $\gamma \leq 0$. As with many other demand manifolds considered in Mrázová and Neary (2017), this implies that, for a given value of σ, the demand manifold has two branches, one in the superconvex region corresponding to negative values of γ, and the other in the subconvex region corresponding to positive values of γ. Along each branch, the equilibrium point converges towards the CES locus as output rises without bound, as shown by the arrows in Figure 4.

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

$$\rho + \varepsilon - 3 = \frac{[(\sigma-2)\varepsilon+1](\varepsilon-1)}{(\sigma-1)\varepsilon} = 0 \tag{48}$$

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 Proofs of Corollaries 1, 2, 3, and 4

Corollaries 1 and 2 (Productivity and Sales with Pareto or Lognormal):

Proposition 2 holds for any distribution in the generalized power function class. The particular solutions for the constant terms in equations (6) and (7) are derived by substituting the parameters of the Pareto and lognormal 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 lognormal distribution, since they imply that the smallest firm has strictly positive sales revenue.

Corollaries 3 and 4 (Productivity and Output with Pareto or Lognormal):

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:

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

where κ is once again a constant of integration. This is the CEMR demand system (8), 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 lognormal.

F Gibrat's Law in Industry Equilibrium

We wish to show that CREMR demands are necessary and sufficient for Gibrat's Law to hold under monopolistic competition following industry-wide productivity shocks. Consider a monopolistically competitive industry with heterogeneous firms. The model is that of Melitz (2003), extended to allow for demands other than CES, following Zhelobodko, Kokovin, Parenti, and Thisse (2012), Bertoletti and Epifani (2014), and Mrázová and Neary (2017). Firms have a common fixed cost f, but differ in their productivity, φ. Each firm produces a unique good, and chooses its level of output y to maximize profits:

$$\pi(\varphi, \lambda, \tau) = \max_{y} \left[\left\{ p(y, \lambda) - \tau \varphi^{-1} \right\} y - f \right] \tag{50}$$

Here, $p(y, \lambda)$ is the inverse demand function faced by all firms, which depends negatively on their output level y and on λ, a common demand parameter that is exogenous to firms but endogenous to the industry. From each firm's perspective, λ is a measure of intensity of competition; we assume that preferences are additively separable, so from the consumer's perspective it equals the marginal utility of income.[31] Finally, τ is a uniform cost shock that is common to all firms.

Firm's productivities are drawn from a distribution $G(\varphi)$ with support on $[\underline{\varphi}, \infty]$. A potential entrant bases its entry decision on the value $v(\varphi, \lambda, \tau)$ that it expects to earn; firm value cannot be negative, so it is zero for firms that get a low-productivity draw and equals operating profits less fixed costs otherwise. Equilibrium requires that the expected value of a firm, $\bar{v}(\lambda, \tau)$, equal the sunk cost of entering the industry f_e:

$$\bar{v}(\lambda, \tau) \equiv \int_{\underline{\varphi}}^{\infty} v(\varphi, \lambda, \tau) g(\varphi) \, \mathrm{d}\varphi = f_e, \qquad \text{where} \qquad v(\varphi, \lambda, \tau) \equiv \max\left[0, \pi(\varphi, \lambda, \tau) - f\right] \tag{51}$$

This zero-expected-profit condition determines the equilibrium value of the intensity of competition λ as a function of the cost parameter τ.

We are now ready to explore conditions for Gibrat's Law to hold following a uniform exogenous chock to the productivity of all firms: $\hat{\tau} < 0$. Writing the sales revenue of a firm of type φ as $r(\varphi) = p(\varphi) y(\varphi)$, the growth rate of sales following a uniform productivity shock is: $g(\varphi) \equiv -\frac{\tau}{r(\varphi)} \frac{dr(\varphi)}{d\tau}$. Hence Gibrat's Law obtains when $g(\varphi)$ is independent of φ: $\frac{dg(\varphi)}{d\varphi} = 0$.

We first consider the effects of the shock on each firm's price and output, taking account

[31] This specification of demand is consistent with other preference systems. The assumption that λ is a scalar implies that the demand functions are members of the generalized separability class of Pollak (1972) already mentioned above; this includes additive separability as well as many other widely-used demand systems.

of the fact that λ is endogenous. Starting with the household's first-order condition, $p(y,\lambda) = \lambda^{-1}u'(\frac{y}{kL})$, we totally differentiate to get the proportional change in prices:

$$\hat{p} = -\frac{1}{\varepsilon}\hat{y} - \hat{\lambda} \tag{52}$$

Hence the change in sales revenue is:

$$\hat{r}(\varphi) = \hat{p}(\varphi) + \hat{y}(\varphi) = \frac{\varepsilon - 1}{\varepsilon}\hat{y}\,(\varphi) - \hat{\lambda} \tag{53}$$

Next we totally differentiate the firm's first-order condition, $p(y,\lambda) + yp_y(y,\lambda) = \tau\varphi^{-1}$, to get the proportional change in outputs:

$$\hat{y} = -\frac{\varepsilon - 1}{2 - \rho}(\hat{\tau} + \hat{\lambda}) \tag{54}$$

Next we totally differentiate the zero-expected-profit condition (51) to get the change in the intensity of competition:

$$\hat{\lambda} = -\frac{\bar{\varepsilon} - 1}{\bar{\varepsilon}}\hat{\tau} \tag{55}$$

Following Mrázová and Neary (2017), $\bar{\varepsilon}$ is the profit-weighted average elasticity of demand across all firms (including those that choose not to enter), which we can interpret as the elasticity faced by the average firm:

$$\bar{\varepsilon} \equiv \int_{\underline{\varphi}}^{\infty} \frac{v(\varphi,\lambda,\tau)}{\bar{v}(\lambda,\tau)}\varepsilon(\varphi)g(\varphi)\,\mathrm{d}\varphi \tag{56}$$

Finally, substituting from (54) and (55) for the changes in y and λ into equation (53), we can calculate the growth rate of a firm with productivity φ following a uniform cost shock ($\hat{\tau} < 0$):

$$g(\varphi) \;\equiv\; -\frac{\hat{r}(\varphi)}{\hat{\tau}} \;=\; \underbrace{\frac{(\varepsilon - 1)^2}{\varepsilon(2-\rho)}}_{(A)}\,\underbrace{\frac{1}{\bar{\varepsilon}} \;-\; \frac{\bar{\varepsilon} - 1}{\bar{\varepsilon}}}_{(B)} \tag{57}$$

To establish whether Gibrat's Law holds, we need to establish when (57) is independent of φ. Clearly the terms indicated by (B) cannot offset the Law irrespective of the form of demand: they are not constant, and will in general be affected by changes in τ, but they are the same for all firms. Hence, differences in growth rates across firms can only arise from differences in (A). It follows that a necessary and sufficient condition for Gibrat's Law in this setting is that (A) is constant across firms. This term, $\frac{(\varepsilon-1)^2}{\varepsilon(2-\rho)}$, is the elasticity of revenue with respect to productivity. It is clearly the inverse of $\frac{\varepsilon(2-\rho)}{(\varepsilon-1)^2}$, which as we have already seen in (46) is the elasticity of marginal revenue with respect to total revenue. We have also seen that this term is constant if and only if demands are CREMR, in which case it equals $\frac{1}{\sigma-1}$. (See Section 3, and equations (44) and (46) in Appendix D.) Hence we have shown that $\frac{dg(\varphi)}{d\tau} = 0$, i.e., Gibrat's Law holds following an industry-wide shock in monopolistic competition, if and only if demands are CREMR.

G Derivations Underlying Table 1

As in Mrázová and Neary (2017), we give the demand functions from a "firm's-eye view"; many of the parameters taken as given by the firm are endogenous in industry and general equilibrium. For each demand function, we follow a similar approach to that used with CREMR demands in Sections 3.1 and 5.2: we use the first-order condition to solve for productivity as a function of either output or price; the definition of sales revenue to solve for output or price as a function of sales; and the relationship between markups and elasticities to solve for 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 The Kullback-Leibler Divergence

H.1 Information-Theoretic Foundations of the KLD

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 $F(r)$.[32] It is natural

[32]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 international trade by Dasgupta and Mondria (2018). Applications of the KLD to economics include Vuong (1989), Cameron and Windmeijer (1997) and Ullah (2002) in econometrics, Adams (2013) in empirical demand analysis, and Galle, Rodríguez-Clare, and Yi (2017) in international trade.

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: $I(r) = -\log(f(r))$.[33] 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:[34]

$$\mathcal{S}_F \equiv E[I(r)] = -\int_{\underline{r}}^{\overline{r}} \log\left(f(r)\right) f(r) dr \tag{58}$$

(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 distribution 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:

$$F(r) = \frac{r - \underline{r}}{\overline{r} - \underline{r}}, \; r \in [\underline{r}, \overline{r}] \quad \Rightarrow \quad \mathcal{S}_F = \mathcal{S}_{Uniform} = \log(\overline{r} - \underline{r}) \tag{59}$$

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 $\tilde{F}(r)$ is the observed c.d.f. of firms' sales, and $F(r)$ is the model-based c.d.f. used to approximate $\tilde{F}(r)$,

[33]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.

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

then the KLD is defined as follows:

$$\mathcal{D}_{KL}\left(\tilde{F} \mid\mid F\right) \equiv \int_{\underline{r}}^{\overline{r}} \log\left(\frac{\tilde{f}(r)}{f(r)}\right) \tilde{f}(r)dr \tag{60}$$

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

$$\mathcal{D}_{KL}\left(\tilde{F} \mid\mid F\right) = -\int_{\underline{r}}^{\overline{r}} \log\left(f(r)\right) \tilde{f}(r)dr - \mathcal{S}_{\tilde{F}} \tag{61}$$

The first term on the right-hand side of (61) measures the cross-entropy between $\tilde{F}(r)$ and $F(r)$. Intuitively, this is a measure of the unpredictability of an individual draw from the benchmark distribution $\tilde{F}(r)$ implied by the tested distribution $F(r)$. Equation (61) 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 $\tilde{F}(r)$ implied by $F(r)$ relative to the unpredictability of $\tilde{F}(r)$ implied by itself; or as the informativeness of a draw from $F(r)$ relative to one from $\tilde{F}(r)$. 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, $\mathcal{D}_{KL}\left(\tilde{F} \mid\mid F\right) \geq 0$, and attains its lower bound of zero if and only if $F(r) = \tilde{F}(r)$ almost everywhere, when the distribution $F(r)$ is completely informative about $\tilde{F}(r)$. As for its upper bound, the KLD value is unbounded unlike Shannon entropy. However, as discussed in the text, we take its value when F is uniform as a benchmark for a "reasonable" fit. This is given by:

$$\mathcal{D}_{KL}\left(\tilde{F} \mid\mid F_{Uniform}\right) = \log(\overline{r} - \underline{r}) - \mathcal{S}_{\tilde{F}} = \mathcal{S}_{Uniform} - \mathcal{S}_{\tilde{F}} \tag{62}$$

where the second equality follows from (60).

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: $\mathcal{D}_{KL}\left(\tilde{F} \mid\mid F\right) \neq \mathcal{D}_{KL}\left(F \mid\mid \tilde{F}\right)$. 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 $F(r)$ approximate the "true" distribution $\tilde{F}(r)$: it measures the divergence of $F(r)$ from $\tilde{F}(r)$, not the distance between them.

Second, for the KLD to be well defined, the tested distribution $F(r)$ must have a strictly positive density, $f(r) > 0$, at every point in $[\underline{r}, \overline{r}]$.[35] In principle, this can pose problems when we wish to compare a distribution implied by a demand function (such as the linear) that implies a saturation consumption level with an unbounded distribution such as the Pareto or lognormal. This is not a problem in practical applications, however, since we can always calibrate demand to fit the upper limit of the observed values of $\tilde{F}(r)$. 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 (2018), 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 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.[36] A more direct approach is to see how the KLD behaves as we

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

[36] 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)): $\mathcal{D}_{\alpha}\big(\tilde{F} \mid\mid F\big) \equiv \frac{1}{\alpha-1}\log\big(\int_{\underline{r}}^{\overline{r}} \frac{f(r)^{\alpha}}{\tilde{f}(r)^{\alpha-1}} dr\big)$. The KLD is the limiting case of this as $\alpha \to 1$: $\mathcal{D}_1\big(\tilde{F} \mid\mid F\big) = \mathcal{D}_{KL}\big(\tilde{F} \mid\mid F\big)$. For values of α between zero and one, the Rényi divergence weights all possible draws more equally than the KLD, regardless of their probability.

drop more observations on smaller firms: we pursue this in Section 7.2.

H.2 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 underlying distributions. To do this we use integration by parts. First, rewrite the definition of Shannon entropy in (58) as $\int_{\underline{r}}^{\overline{r}} u dv$, 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:

$$S_F = -(1 + C) \log f(\overline{r}) + C \log f(\underline{r}) + \int_{\underline{r}}^{\overline{r}} \frac{F(r) + C}{r} \frac{r f'(r)}{f(r)} dr \tag{63}$$

Setting C equal to -1 gives:

$$S_F = -\log f(\underline{r}) - \int_{\underline{r}}^{\overline{r}} \frac{1 - F(r)}{r} \frac{r f'(r)}{f(r)} dr \tag{64}$$

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.[37] If instead we set C in (63) equal to zero rather than one, we get an alternative decomposition expressed in terms of the upper bound:

$$S_F = -\log f(\overline{r}) + \int_{\underline{r}}^{\overline{r}} \frac{F(r)}{r} \frac{r f'(r)}{f(r)} dr \tag{65}$$

[37] The rate at which the relative survival function declines is one plus the proportional hazard rate: $d \log \left[\frac{1-F(r)}{r}\right] = -\left(1 + \frac{rf}{1-F}\right) d \log r$.

Now the first term is the information content of the upper limit of the distribution. However, this is less useful than (64) for our purposes, since, for many distributions, including the Pareto and the lognormal, $\log f(\overline{r}) = -\infty$.

Repeating this process for the KLD gives in a similar fashion two alternative decompositions, one expressed in terms of the lower bounds of the distribution:

$$\mathcal{D}_{KL}\left(\tilde{F} \mid\mid F\right) = \log \tilde{f}\left(\underline{r}\right) - \log f\left(\underline{r}\right) + \int_{\underline{r}}^{\overline{r}} \frac{1-\tilde{F}\left(r\right)}{r} \left[\frac{r\tilde{f}'\left(r\right)}{\tilde{f}\left(r\right)} - \frac{rf'\left(r\right)}{f\left(r\right)}\right] dr \tag{66}$$

and the other in terms of the upper bounds:

$$\mathcal{D}_{KL}\left(\tilde{F} \mid\mid F\right) = \log \tilde{f}\left(\overline{r}\right) - \log f\left(\overline{r}\right) - \int_{\underline{r}}^{\overline{r}} \frac{\tilde{F}\left(r\right)}{r} \left[\frac{r\tilde{f}'\left(r\right)}{f\left(r\right)} - \frac{rf'\left(r\right)}{f\left(r\right)}\right] dr \tag{67}$$

Again we concentrate on the first of these, which, as before, 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, $\frac{1-\tilde{F}(r)}{r}$. 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 (66) 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: $f(r) = g(\varphi(r))\frac{d\varphi}{dr}$. Totally differentiating this gives an expression in terms of elasticities:

$$\frac{rf'(r)}{f(r)} = \frac{\varphi(r)g'(\varphi(r))}{g(\varphi(r))}\frac{r\varphi'(r)}{\varphi(r)} + \frac{r\varphi''(r)}{\varphi'(r)} \tag{68}$$

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

$$\frac{r\varphi''(r)}{\varphi'(r)} = E(r) - 1 + \frac{rE'(r)}{E(r)} \tag{69}$$

(See Lemma 5 in Mrázová and Neary (2017).) Substituting into (68), the density elasticity of the derived sales distribution $F(r)$ can be written in terms of underlying elasticities as follows:

$$\frac{rf'(r)}{f(r)} = \left[\frac{\varphi(r)g'(\varphi(r))}{g(\varphi(r))} + 1\right] E(r) - 1 + \frac{rE'(r)}{E(r)} \tag{70}$$

Substituting this into (66) gives the full decomposition of the KLD in equation (35) in the text. When $G(\varphi)$ is Pareto, so $G(\varphi) = 1 - \left(\frac{\varphi}{\underline{\varphi}}\right)^{-k}$, the elasticity of density is $\frac{\varphi g'(\varphi)}{g(\varphi)} = -(1+k)$. Hence (70) simplifies to the following:

$$\frac{rf'(r)}{f(r)} = -[kE(r) + 1] + \frac{rE'(r)}{E(r)} \tag{71}$$

H.3 Quantifying the Information Cost of Incorrect Assumptions

To illustrate the application of the KLD decomposition in equation (35), 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 (35), 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 respectively, the KLD can be calculated from equation (66):

$$\mathcal{D}_{KL}(\tilde{F} \,||\, F) = \log \frac{\tilde{n}}{n} + \frac{n}{\tilde{n}} - 1 \tag{72}$$

To relate this to primitive parameters, 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

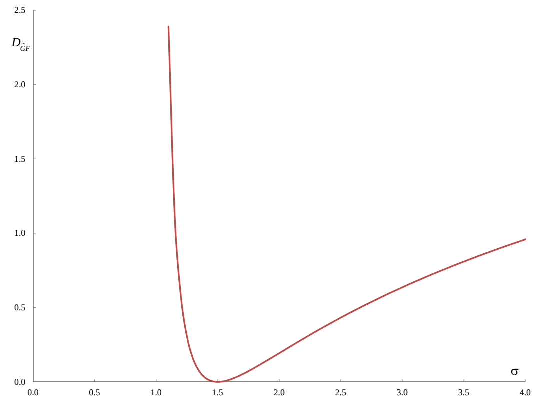


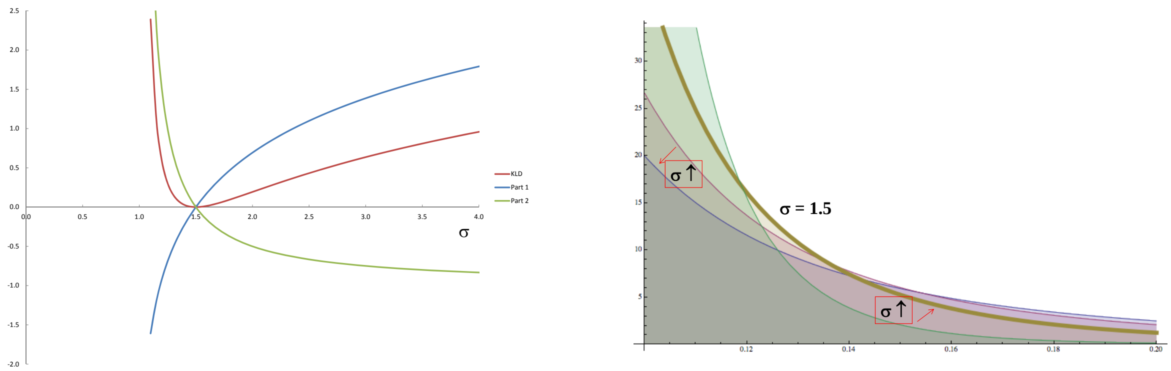
Figure 12: KLD as a Function of σ in the Pareto-CREMR Case

Pareto distribution of productivity, the shape parameter for the derived distribution of sales is $n = Ek = \frac{k}{\sigma-1}$. Substituting into (72) gives the KLD decomposition, equation (35), in the Pareto-CREMR case:

$$\mathcal{D}_{KL}(\tilde{F} \mid\mid F) = \underbrace{\log \frac{\tilde{n}}{k} + \log(\sigma - 1)}_{(1)} + \underbrace{\frac{k}{\tilde{n}} \frac{1}{\sigma - 1} - 1}_{(2)} \tag{73}$$

The first derivative of this with respect to σ is: $\frac{d\mathcal{D}_{KL}}{d\sigma} = (\sigma-1)^{-1}\left(1 - \frac{n}{\tilde{n}}\right) = (\sigma-1)^{-2}\left(\sigma - \frac{k+\tilde{n}}{\tilde{n}}\right)$. This is positive, and so $\mathcal{D}_{KL}$ is increasing, if and only if $\sigma \geq \frac{k+\tilde{n}}{\tilde{n}}$. The second derivative is: $\frac{d^2\mathcal{D}_{KL}}{d\sigma^2} = -(\sigma-1)^{-2}\left(1 - 2\frac{n}{\tilde{n}}\right) = -(\sigma-1)^{-3}\left(\sigma - \frac{2k+\tilde{n}}{\tilde{n}}\right)$. This is negative, and so $\mathcal{D}_{KL}$ is concave, if and only if $\sigma \geq \frac{2k+\tilde{n}}{\tilde{n}}$. Equation (73) is illustrated in Figure 12 as a function of σ, drawn for values of $k = 1$ and $\tilde{n} = 2$.

Figure 12 shows clearly that the information cost of using the "wrong" estimate of σ is highly asymmetric. For given values of k and $\tilde{n}$, the true value of σ equals $\frac{k+\tilde{n}}{\tilde{n}}$. (Recall equation (6).) Given the assumed values of k and $\tilde{n}$, 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 13. Panel (a) shows the two numbered components of the KLD from (73), while panel (b) shows how a higher assumed value of σ affects the location of



(a) Components of KLD (b) Effects of Changes in σ on Fitted Pareto

Figure 13: Decomposition of KLD in the Pareto-CREMR Case

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 (73), 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 (73) is that, with Pareto productivity and CREMR demands, the KLD depends on only one of the three parameters in the CREMR demand function. Figure 12 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.

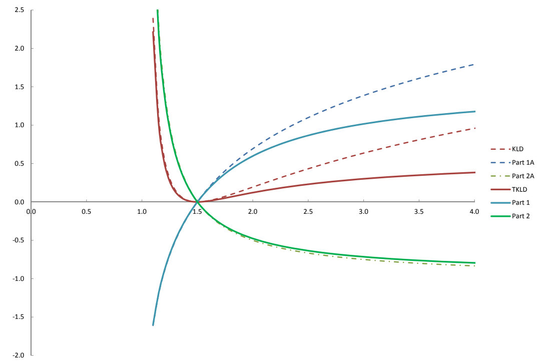


Figure 14: Components of KLD in the Truncated-Pareto-CREMR Case

It is also illuminating to see how the KLD extends to the case of CREMR demands combined with productivity and observed sales distributions that are right-truncated Paretos, $F(r) = \frac{1-\underline{r}^n r^{-n}}{1-\underline{r}^n \overline{r}^{-n}}$ and $\tilde{F}(r) = \frac{1-\underline{r}^{\tilde{n}} r^{-\tilde{n}}}{1-\underline{r}^{\tilde{n}} \overline{r}^{-\tilde{n}}}$, where $r \in [\underline{r}, \overline{r}]$. In the untruncated Pareto case, the KLD was independent of the lower bound of the Pareto $\underline{r}$. This is no longer true, though it depends only on the ratio of the lower and upper bounds: $\lambda \equiv \frac{\underline{r}}{\overline{r}} \in [0, 1]$. (This reduces to zero in the untruncated case.) Straightforward calculations give the extension of equation (72) to the truncated case:

$$\mathcal{D}_{KL}(\tilde{F} \,||\, F) = \log \frac{\tilde{n}}{n} + \log \frac{1-\lambda^n}{1-\lambda^{\tilde{n}}} + \frac{n-\tilde{n}}{\tilde{n}} + (n-\tilde{n}) \frac{\lambda^{\tilde{n}}}{1-\lambda^{\tilde{n}}} \log \lambda \tag{74}$$

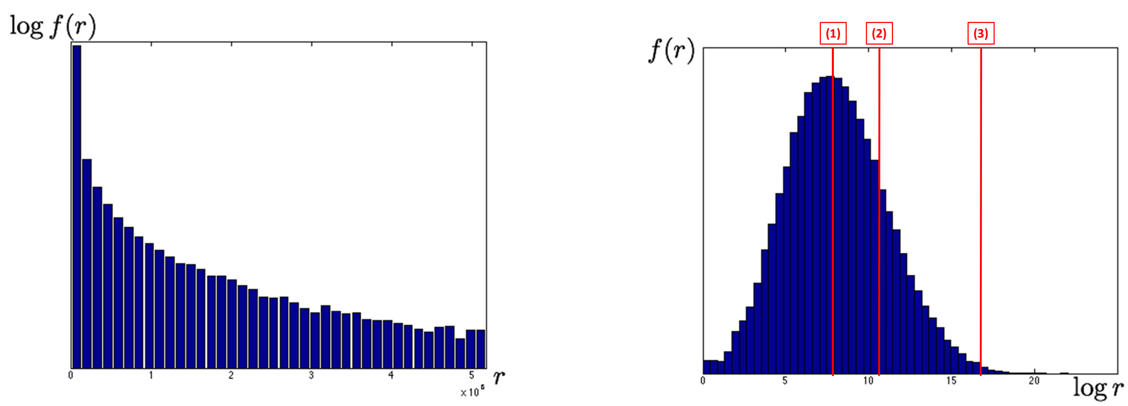
Replacing some, though not all, occurrences of n by $\frac{k}{\sigma-1}$ allows us to write this in terms of primitives in a manner which parallels equation (73):

$$\mathcal{D}_{KL}(\tilde{F} \,||\, F) = \underbrace{\log \frac{\tilde{n}}{k} + \log(\sigma - 1) + \log \frac{1-\lambda^n}{1-\lambda^{\tilde{n}}}}_{(1)} + \underbrace{\frac{k}{\tilde{n}} \frac{1}{\sigma - 1} - 1 + (n-\tilde{n}) \frac{\lambda^{\tilde{n}}}{1-\lambda^{\tilde{n}}} \log \lambda}_{(2)} \tag{75}$$

This is illustrated in Figure 14 as a function of σ. The dashed loci repeat the KLD and its components for the untruncated case from Figure 13(a). The solid loci give the KLD and its components for the truncated case, assuming the same values of k and $\tilde{n}$ as before and a value of $\lambda = 0.1$. It is clear that right-truncation makes almost no quantitative difference to the cost of underestimating σ. The main effect is to reduce the cost of mismeasuring the mass of the smallest firms when σ is overestimated.

I French Exports to Germany

The Indian data used in Section 7 have the great advantage that they give both sales and markups for all firms. This is important, for example, in allowing us to discriminate between CES and CREMR, whose implications for sales are observationally equivalent. However, relative to many data sets used in recent trade applications, they refer to total sales rather than exports and they cover a relatively small number of firms. Hence it is useful to repeat the analysis on a more conventional data set on export sales, even if this does not give information on markups. We do this in this section, using data on the universe of French exports to Germany in 2005, drawn from the same source as that used by Head, Mayer, and Thoenig (2014).[38]



(a) A First Look: Obviously Pareto?

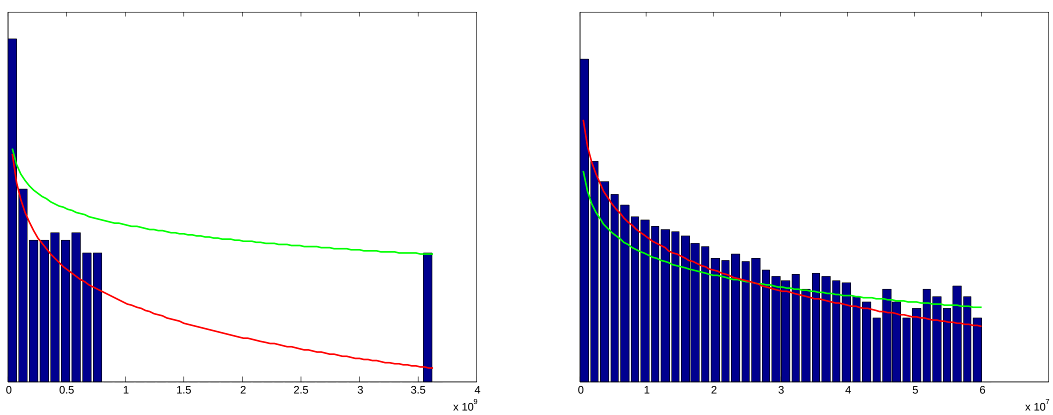
(b) A Second Look: Obviously Lognormal?

Figure 15: Alternative Perspectives on the Data

Figure 15 shows that the export sales data exhibit some typical features of such data

[38]The data set contains 161,191 firm-product observations on export sales by 27,550 firms: 5.85 products per firm. We are very grateful to Julien Martin for performing the analysis for us on French Customs 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 lognormal. Yet a 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) All Observations (b) All Bar Top 89 Observations

Figure 16: KLD-Minimizing Predicted Distributions: Pareto (green) and Lognormal (red)

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 16 compares the best-fit Pareto (in green) and lognormal (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 lognormal. (Recall from our discussion in the text that the distribution and demand parameters are not separately identified.) Panel (a) illustrates the results for all firms, while panel (b) avoids the distorted perspective caused by including the largest firms, by omitting the top 89 firms, which account for 0.05% of observations but 32% of sales. Inspecting the fitted distributions, it is evident that the lognormal matches the smaller firms better, and conversely for the Pareto. The values of the minimized KLD show that the lognormal provides a better overall fit than the Pareto: 0.0001 as opposed to 0.0012. (As with the Indian data in Section 7, the data are normalized by the value of the KLD for a uniform distribution, which for this data set is 6.8082.)

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

Table 5: KLD for French Exports Compared with Predictions from Selected Demand Functions and Productivity Distributions

Table 5 gives the values of the KLD for the Pareto and lognormal cases shown in Figure 16, and also for the distributions implied by either translog or linear demand functions combined with either Pareto or lognormal productivities. These distributions are calculated by combining the relevant productivity distribution with the relationships between productivity and sales implied by translog and linear demands from Table 1. (Recall from that table that the linear and LES specifications are observationally equivalent.) Each entry in the table is the value of the KLD that measures the information loss when the combination of assumptions indicated by the row and column is used to explain the observed distribution of sales.

To assess whether the values are significantly different from one another, we use a bootstrapping approach. We construct one thousand samples of the same size as the data (i.e., 161,191 observations), by sampling with replacement from the original data. For each sample, we then compute the KLD value for each of the six models. Table 6 gives the results. Each entry in the table is the proportion of samples in which the combination in the relevant column gives a higher value of the KLD than that in the relevant row. All the values are equal to or very close to 100%, which confirms that the results in Table 5 are robust.

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

Table 6: Bootstrapped Robustness of the KLD Ranking: French Sales
(See text for explanation)

Turning to the results in Table 5, recall that panel (a) of Figure 16 showed that the lognormal matches the smaller firms better, and conversely for the Pareto. Table 5 provides a quantitative confirmation of this. With a preponderance of the bins corresponding to smaller firms, it is not surprising that the lognormal does better as measured by the KLD. However, the difference between distributions turns out to be much less significant than that between different specifications of demand. The KLD values for the translog/AI and linear/LES specifications are much higher than for the CREMR case, as shown in the third and fourth columns of Table 5, with the Pareto now preferred to the lognormal. The overwhelming conclusion from these results is that, if we want to fit the distribution of sales in this data set, then the choice between Pareto and lognormal distributions is less important than the choice between CREMR and other demands. This is broadly in line with the results for Indian sales data in Section 7, especially when we exclude the smallest firms as in Section 7.2.

J Bootstrapped Comparisons on Indian Data

Tables 7 and 8 repeat for Indian sales and markup data respectively the bootstrapping comparisons presented for French exports data in Table 6. It is clear that the comparisons between different values of the KLD for Indian data shown in Table 2 and Figure 8 in the text are just as robust as those for the French data shown in Table 5.

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

Table 7: Bootstrapped Robustness of the KLD Ranking: Indian Sales
(See text for explanation)

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

Table 8: Bootstrapped Robustness of the KLD Ranking: Indian Markups
(See text for explanation)

K Quantiles for the QQ Estimator

Tables 9 and 10 give the expressions for the quantiles of the sales and markup distributions respectively that are implied by our assumptions about demand and the distribution of firm productivities. These expressions are used to calculate the entries in Figure 11.

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

Table 9: Quantiles for Sales

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

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

Table 10: Quantiles for Markups

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

References

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

Antràs, P., and D. Chor (2013): "Organizing the Global Value Chain," *Econometrica*, 81(6), 2127–2204.

Antràs, P., and E. Helpman (2004): "Global Sourcing," *Journal of Political Economy*, 112(3), 552–580.

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

——— (2010b): "Market Penetration Costs and the New Consumers Margin in International Trade: Appendix," web appendix, Yale University.

——— (2016): "A Unified Theory of Firm Selection and Growth," *Quarterly Journal of Economics*, 131(1), 89–155.

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

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.

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293(5536), 1818–1820.

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

Bee, M., and S. Schiavo (2018): "Powerless: Gains from Trade when Firm Productivity is not Pareto Distributed," *Review of World Economics*, 154(1), 15–45.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014): "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 122(3), 507–553.

Bernard, A. B., S. J. Redding, and P. K. Schott (2011): "Multi-Product Firms and Trade Liberalization," *Quarterly Journal of Economics*, 126(3), 1271–1318.

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

Bustos, P. (2011): "Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American Economic Review*, 101(1), 304–340.

Cabral, L. M., and J. Mata (2003): "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 93(4), 1075–1090.

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.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection," *Econometrica*, 80(6), 2543–2594.

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

Dasgupta, K., and J. Mondria (2018): "Inattentive Importers," *Journal of International Economics*, 112, 150–165.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, Markups and Trade Reform," *Econometrica*, 84(2), 445–510.

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

DHINGRA, S., AND J. MORROW (2019): "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," forthcoming, *Journal of Political Economy*.

DI GIOVANNI, J., AND A. A. LEVCHENKO (2012): "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies," *Journal of Political Economy*, 120(6), 1083–1132.

DIXIT, A. K., AND J. E. STIGLITZ (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3), 297–308.

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

EECKHOUT, J. (2004): "Gibrat's Law for (All) Cities," *American Economic Review*, 94(5), 1429–1451.

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.

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

FEENSTRA, R., AND H. L. KEE (2008): "Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity," *Journal of International Economics*, 74(2), 500–518.

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

Feenstra, R. C., and D. E. Weinstein (2017): "Globalization, Markups, and U.S. Welfare," *Journal of Political Economy*, 125(4), 1040–1074.

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

——— (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79(3), 733–772.

Galle, S., A. Rodríguez-Clare, and M. Yi (2017): "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade," NBER Working Paper No. 23737.

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

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.

Hopenhayn, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5), 1127–1150.

Hsieh, C.-T., and P. J. Klenow (2007): "Relative Prices and Relative Prosperity," *American Economic Review*, 97(3), 562–585.

Ijiri, Y., and H. A. Simon (1974): "Interpretations of Departures from the Pareto Curve Firm-Size Distributions," *Journal of Political Economy*, 82(2), 315–331.

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,” *International Statistical Review*, 83(2), 175–192.

Jovanovic, B. (1982): “Selection and the Evolution of Industry,” *Econometrica*, 50(3), 649–670.

Kimball, M. S. (1990): “Precautionary Saving in the Small and in the Large,” *Econometrica*, 58(1), 53–73.

Kratz, M., and S. I. Resnick (1996): “The QQ-estimator and heavy tails,” *Stochastic Models*, 12(4), 699–724.

Krusell, P., and A. A. Smith, Jr. (1998): “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 106(5), 867–896.

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.

Luttmer, E. G. (2007): “Selection, Growth, and the Size Distribution of Firms,” *Quarterly Journal of Economics*, 122(3), 1103–1144.

——— (2010): “Models of Growth and Firm Heterogeneity,” *Annual Reviews of Economics*, 2(1), 547–576.

——— (2011): “On the Mechanics of Firm Growth,” *Review of Economic Studies*, 78(3), 1042–1068.

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.

Melitz, M. J., and S. J. Redding (2015): "New Trade Models, New Welfare Implications," *American Economic Review*, 105(3), 1105–46.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

Mrázová, M., and J. P. Neary (2017): "Not so Demanding: Demand Structure and Firm Behavior," *American Economic Review*, 107(12), 3835–3874.

——— (2019): "Selection Effects with Heterogeneous Firms," forthcoming in *Journal of the European Economic Association*.

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

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

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

Pollak, R. A. (1972): "Generalized Separability," *Econometrica*, 40(3), 431–453.

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. (2015): "Income Differences and Prices of Tradables: Insights from an Online Retailer," *Review of Economic Studies*, 82(4), 1612–1656.

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

Stiglitz, J. E. (1969): "Distribution of Income and Wealth Among Individuals," *Econometrica*, 37(3), 382–397.

Sutton, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35(1), 40–59.

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. (2017): "Global Production with Export Platforms," *Quarterly Journal of Economics*, 132(1), 157–209.

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.