

SUPPLEMENT TO “SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS”

(*Econometrica*, Vol. 89, No. 4, July 2021, 1753–1788)

MONIKA MRÁZOVÁ

Geneva School of Economics and Management, University of Geneva, CEPR, and CESifo

J. PETER NEARY

Department of Economics, University of Oxford, CEPR, and CESifo

MATHIEU PARENTI

ECARES, Université Libre de Bruxelles and CEPR

APPENDIX B: SUPPLEMENTAL ONLINE APPENDIX

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

B.1. *Other Implications of Proposition 1*

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

B.1.1. *Self-Reflection of Productivity and Output*

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

$$p(x) = \frac{1}{x}(\alpha + \beta x^{\frac{\sigma-1}{\sigma}}). \quad (35)$$

This demand function plays the same role with respect to firm output as the CREMR demand function does with respect to firm sales (recall (6)). It is necessary and sufficient for a constant elasticity of marginal revenue with respect to output, equal to $E = \frac{1}{\sigma}$. Hence, we call it “CEMR” for “Constant (Output) Elasticity of Marginal Revenue.”²⁶

Monika Mrázová: monika.mrazova@unige.ch

J. Peter Neary: (Deceased)

Mathieu Parenti: mathieu.parenti@ulb.ac.be

²⁶“CEMR” rhymes with “seemer,” just as “CREMR” rhymes with “dreamer.”

Unlike CREMR, there are some precedents for the CEMR class. It has the same functional form, except with prices and quantities reversed, as the direct PIGL (“Price-Independent Generalized Linearity”) class of [Muellbauer \(1975\)](#).²⁷ In particular, the limiting case where σ approaches 1 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.²⁸ When the common distribution of productivity and output is a Pareto, we can immediately state a further corollary of Proposition 1:

COROLLARY 4: *Given Assumption 1, any two of the following imply the third:*

(A) *The distribution of firm productivity is Pareto: $G_{\mathcal{P}}(\varphi) = 1 - \varphi^k \varphi^{-k}$.*

(B) *The distribution of firm output is Pareto: $F_{\mathcal{P}}(x) = 1 - \underline{x}^m x^{-m}$.*

(C) *The demand function belongs to the CEMR family in (35),*

where the parameters are related as follows:

$$m = \frac{k}{\sigma} \quad \text{and} \quad \underline{x} = \left(\beta \frac{\sigma - 1}{\sigma} \varphi \right)^{\sigma}.$$

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

COROLLARY 5: *Given Assumption 1, any two of the following imply the third:*

(A) *The distribution of firm productivity is truncated lognormal with support $[\underline{\varphi}, +\infty)$:*

$$G_{tLN}(\varphi) = \frac{\Phi((\log \varphi - \mu)/s) - T}{1 - T}.$$

(B) *The distribution of firm output is truncated lognormal with support $[\underline{x}, +\infty)$:*

$$F_{tLN}(x) = \frac{\Phi((\log x - \mu'')/s'') - T}{1 - T}.$$

(C) *The demand function belongs to the CEMR family in (35),*

where the parameters are related as follows:

$$s'' = \sigma s,$$

$$\mu'' = \sigma \left(\mu + \log \left(\beta \frac{\sigma - 1}{\sigma} \right) \right),$$

$$\underline{x} = \left(\beta \frac{\sigma - 1}{\sigma} \varphi \right)^{\sigma},$$

$$T = \Phi((\log \varphi - \mu)/s) = \Phi((\log \underline{x} - \mu'')/s'').$$

B.1.2. Self-Reflection of Output and Sales

A final self-reflection corollary of Proposition 1 relates to the case where output and sales follow the same distribution. This requires that the elasticity of one with respect to

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

²⁸As shown by [Mrázová and Neary \(2017\)](#), the CEMR demand manifold implies a linear relationship between the convexity and elasticity of demand, passing through the Cobb–Douglas point $(\varepsilon, \rho) = (1, 2)$: $\rho = 2 - \frac{\varepsilon - 1}{\sigma}$. The manifold for the inverse translog special case ($\sigma \rightarrow 1$) coincides with the SM locus in Figure 2(b). For smaller firms when demand is subconvex, CEMR demands are qualitatively similar to CREMR, except that they are somewhat more elastic: the CREMR manifold from (11) is, for large ε , asymptotically equivalent to $\rho = 2 - \frac{\varepsilon - 2}{\sigma - 1}$.

the other is constant, which implies that the demand function must be a CES.²⁹ Formally, we have the following:

COROLLARY 6: *Given Assumption 1, any two of the following imply the third:*

- (A) *The distribution of firm output x is a member of the generalized power function class.*
- (B) *The distribution of firm sales revenue r is a member of the same family of the generalized power function class.*
- (C) *The demand function is CES: $p(x) = \beta x^{-\frac{1}{\sigma}}$, where $\beta = x_0^{-\frac{1}{\sigma}}$ and $\sigma = \frac{E}{E-1}$.*

In the Pareto case, the sufficiency part of this result is familiar from the large literature on the Melitz model with CES demands: it is implicit in [Chaney \(2008\)](#), for example. The necessity part, taken together with earlier results, shows that it is not possible for all three firm attributes, productivity, sales, and revenue, to have the same distribution from the generalized power class under any demand system other than the CES. Corollary 6 follows immediately from previous results when productivities themselves have a generalized power function distribution, since the only demand function which is a member of both the CEMR and CREMR families is the CES itself. However, it is much more general than that, since it does not require any assumption about the underlying distribution of productivities. It is an example of a corollary to Proposition 1 which relates two endogenous firm outcomes rather than an exogenous and an endogenous one.

B.2. Further Properties of CREMR Demand Functions

To establish conditions for demand to be superconvex, we solve for the points of intersection between the demand manifold and the CES locus, the boundary between the sub- and superconvex regions. From [Mrázová and Neary \(2017\)](#), the expression for the CES locus is $\rho = \frac{\varepsilon+1}{\varepsilon}$. Eliminating ρ using the CREMR demand manifold (11) and factorizing gives

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

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 1, we find that CREMR demands are superconvex if and only if $\gamma \leq 0$. As with many other demand manifolds considered in [Mrázová and Neary \(2017\)](#), this implies that, for a given value of σ , the demand manifold has two branches, one in the superconvex region corresponding to negative values of γ , and the other in the subconvex region corresponding to positive values of γ . Along each branch, the equilibrium point converges towards the CES locus as output rises without bound, as shown by the arrows in Figure 2.

Similarly, to establish conditions for profits to be supermodular, we solve for the points of intersection between the demand manifold and the SM locus, the boundary between

²⁹Suppose 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.

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

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

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

B.3. *Fitting Sales and Markup Distributions*

Section 6.2 in the text focused on how different assumptions compare in predicting the distribution of markups. Here we supplement this by showing in addition how they compare in predicting the distribution of sales. To compare the “goodness of fit” of different models, we use the Kullback–Leibler Divergence (denoted “KLD” hereafter), introduced by [Kullback and Leibler \(1951\)](#). This measures the divergence of the predicted distribution from the actual one, and is asymptotically equivalent to maximum likelihood.³⁰ It equals the information loss from using the theory rather than knowing the true distribution. Whereas information scientists typically present KLD values in “bits” (log to base 2) or “nats” (log to base e), units with little intuitive appeal in economics, we present its values normalized by the value implied by a uniform distribution. This is an uninformative prior in the spirit of the Laplace principle of insufficient reason; it is analogous to the “dartboard” approach to benchmarking the geographic concentration of manufacturing industry of [Ellison and Glaeser \(1997\)](#), or the “balls and bins” approach to benchmarking the world trade matrix of [Armenter and Koren \(2014\)](#). The value of the KLD is unbounded, but a specification that gave a value greater than that implied by a uniform distribution would be an unsatisfactory explanation of the data. Appendix B.5 shows that an alternative criterion for choosing between distributions, the QQ estimator, gives qualitatively similar results.

The minimized KLD values for each specification are given in Table B.I and illustrated in Figure B.1. The rankings of different specifications for sales are very different in the Pareto and lognormal cases. Conditional on a Pareto distribution of productivities, CREMR demands give the worst fit to sales, with translog demands performing best, and linear-LES intermediate between the others. However, the differences between the KLD values for these specifications are much less than those conditional on lognormal productivities. In this case, CREMR does best, with translog performing much less well and linear-LES worst of all.

As for the results for markups, these imply exactly the same ranking of different specifications as the estimates given in Table II in Section 6.2, despite the different methodologies used (using individual observations and minimizing the AIC rather than using data grouped in bins and minimizing the KLD as here). Once again, CREMR demands clearly do best, irrespective of the assumed distribution, with translog and LES performing at the

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

TABLE B.I

KLD FOR INDIAN SALES AND MARKUPS COMPARED WITH PREDICTIONS FROM SELECTED PRODUCTIVITY DISTRIBUTIONS AND DEMAND FUNCTIONS^a

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

^aEach KLD value measures the divergence of the predicted from the empirical distribution. A value of zero indicates no divergence, a value of 1 a divergence as great as a uniform distribution.

same level, and linear doing better under Pareto assumptions but less well in the lognormal case. This reinforces the conclusion drawn in the text that the choice between Pareto and lognormal distributions is less important than the choice between CREMR and other demands.

To assess whether the KLD values in Table B.I 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., 2457 observations), by sampling with replacement from the original data. For each sample, we then compute the KLD value for each of the six models. Tables B.II and B.III give the results for Indian sales and markup data, respectively. Each entry in the table is the proportion of samples in which the combination in the relevant column gives a higher value of the KLD than that in the relevant row. All the values are equal to or very close to 100%, which confirms that the results in Table B.I are robust.

B.4. Robustness to Truncation

The results for Indian sales data in the preceding subsection are broadly similar to those with French sales data in Appendix B.6 below, except for the case of CREMR demands combined with Pareto productivity: this gives a good fit with French data but performs

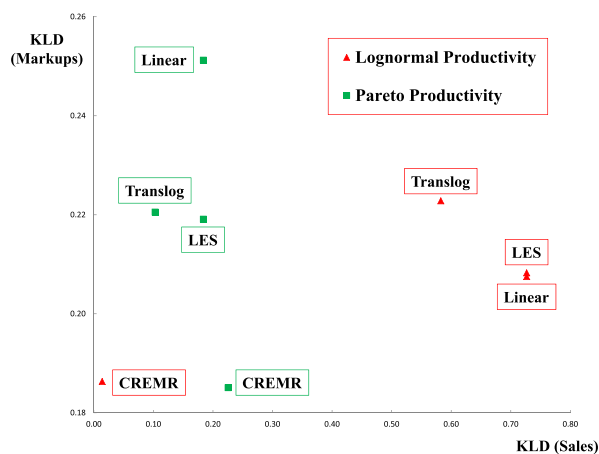


FIGURE B.1.—KLD for Indian sales and markups. Data from Table B.I.

TABLE B.II
BOOTSTRAPPED ROBUSTNESS OF THE KLD RANKING: INDIAN SALES^a

	CREMR + \mathcal{LN}	CREMR + \mathcal{P}	Translog + \mathcal{P}	Linear + \mathcal{P}	Translog + \mathcal{LN}	Linear + \mathcal{LN}
CREMR + \mathcal{LN}	–	0%	0%	0%	0%	0%
CREMR + \mathcal{P}	100%	–	0%	0%	0%	0%
Translog + \mathcal{P}	100%	100%	–	0%	0%	0%
Linear + \mathcal{P}	100%	100%	100%	–	0%	0%
Translog + \mathcal{LN}	100%	100%	100%	100%	–	0%
Linear + \mathcal{LN}	100%	100%	100%	100%	100%	–

^aSee text for explanation. “ \mathcal{LN} ” denotes lognormal, “ \mathcal{P} ” denotes Pareto.

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 B.2 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 B.1) and successively dropping up to 809 observations one at a time.³¹ Although the curves are not precisely monotonic, the broad picture is clear: conditional on CREMR demands, Pareto does better and lognormal does worse as more and more observations are 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 B.3 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 B.2.) In this case, the CREMR specification overtakes the linear one when we drop

TABLE B.III
BOOTSTRAPPED ROBUSTNESS OF THE KLD RANKING: INDIAN MARKUPS^a

	CREMR + \mathcal{P}	CREMR + \mathcal{LN}	Linear + \mathcal{LN}	LES + \mathcal{LN}	LES + \mathcal{P}	Translog + \mathcal{P}	Translog + \mathcal{LN}	Linear + \mathcal{P}
CREMR + \mathcal{P}	–	2%	0%	0%	0%	0%	0%	0%
CREMR + \mathcal{LN}	98%	–	0%	0%	0%	0%	0%	0%
Linear + \mathcal{LN}	100%	100%	–	0%	0%	0%	0%	0%
LES + \mathcal{LN}	100%	100%	100%	–	0%	0%	0%	0%
LES + \mathcal{P}	100%	100%	100%	100%	–	16%	6%	0%
Translog + \mathcal{P}	100%	100%	100%	100%	84%	–	0%	0%
Translog + \mathcal{LN}	100%	100%	100%	100%	94%	100%	–	0%
Linear + \mathcal{P}	100%	100%	100%	100%	100%	100%	100%	–

^aSee text for explanation. “ \mathcal{LN} ” denotes lognormal, “ \mathcal{P} ” denotes Pareto.

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

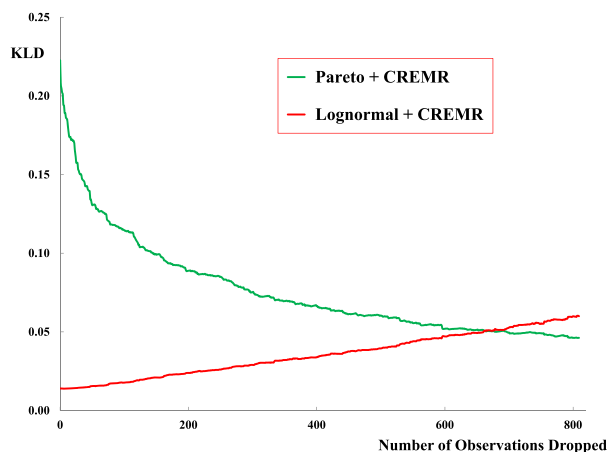


FIGURE B.2.—CREMR versus CREMR: KLD for Indian sales.

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

These findings confirm that the combination of CREMR demand and Pareto productivities fits the sales data relatively better when the smallest observations are dropped. They also make precise the pattern observed in many data sets, whereby the Pareto assumption outperforms the lognormal in the right tail of the sales distribution. For example, Figure B.2 shows that the relevant region in the right tail begins at exactly 663 observations.

B.5. Robustness to Divergence Criterion: The QQ Estimator

To check the robustness of our results, we consider an alternative criterion to the KLD for comparing predicted and actual distributions. Here we consider the QQ estimator, developed by [Kratz and Resnick \(1996\)](#), and previously used by [Head, Mayer, and Thoenig](#)

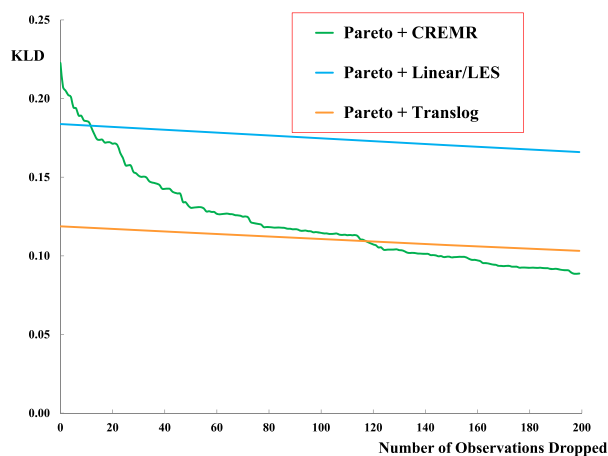


FIGURE B.3.—CREMR versus the rest, given Pareto: KLD for Indian sales.

(2014) and Nigai (2017). This estimator does not have the same desirable theoretical properties as the KLD, in particular it is not asymptotically equivalent to maximum likelihood, but it has a simple interpretation. It equals the parameter vector θ^* that minimizes the sum of the squared deviations of the quantiles of the predicted distribution from those of the actual distribution:

$$QQ(\tilde{F}\|F(\cdot; \theta)) = \sum_{i=1}^n (\log \tilde{q}_i - \log q_i(\theta))^2,$$

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

To implement the QQ estimator, we need analytic expressions for the quantiles of the sales and markup distributions under each of the eight combinations of assumptions about demand and the distribution of productivity we consider. These are given in Tables B.IV and B.V. 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 B.VI, and they are illustrated in Figure B.4.

Comparing Table B.VI and Figure B.4 with Table B.I and Figure B.1 in Appendix B.3 respectively, it is evident that the results based on the QQ estimator are qualitatively very similar to those for the KLD. In particular, the Pareto assumption gives a better fit for sales than for markups, except in the CREMR case; while the lognormal assumption tends to give a better fit for markups than for sales. Comparing different demand functions, CREMR demands give a better fit to the markup distribution than any other demands, irrespective of which productivity distribution is assumed. As for sales, the results differ between the Pareto and lognormal cases. Conditional on lognormal, CREMR again performs much better, whereas, conditional on Pareto, it performs least well, with the translog doing best. The only 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 in Section 6.1 are not unduly sensitive to our choice of criterion for comparing actual and predicted distributions.

TABLE B.IV

QUANTILE FUNCTIONS FOR SALES DISTRIBUTIONS IMPLIED BY ASSUMPTIONS ABOUT PRODUCTIVITY (PARETO (\mathcal{P}) OR TRUNCATED LOGNORMAL ($t\mathcal{LN}$)) AND DEMAND (CREMR, LINEAR, LES, OR TRANSLOG)^a

Demand Function	Pareto Productivity $G_{\mathcal{P}}(\varphi) = 1 - \underline{\varphi}^k \varphi^{-k}$	Truncated Lognormal Productivity $G_{t\mathcal{LN}}(\varphi) = \frac{\Phi(\frac{\log \varphi - \mu}{\sigma}) - T}{1 - T}$
CREMR $p(x) = \frac{\beta}{x}(x - \gamma)^{\frac{\sigma-1}{\sigma}}$	$Q(y) = \beta \sigma \left(\frac{\sigma-1}{\sigma} \underline{\varphi}\right)^{\sigma-1} (1-y)^{\frac{1-\sigma}{k}}$	$Q(y) = \beta \sigma \left(\frac{\sigma-1}{\sigma}\right)^{\mu+s\Phi^{-1}(y(1-T)+T)} \sigma^{-1}$
Linear $p(x) = \alpha - \beta x$	$Q(y) = \frac{\alpha^2 \varphi^2 - (1-y)^{\frac{2}{k}}}{4\beta \varphi^2}$	$Q(y) = \frac{\alpha^2 - e^{-2(\mu+s\Phi^{-1}(y(1-T)+T))}}{4\beta}$
LES $p(x) = \frac{\delta}{x+\gamma}$	$Q(y) = \delta - \sqrt{\frac{\gamma\delta}{\varphi}} (1-y)^{\frac{1}{2k}}$	$Q(y) = \delta - \frac{\sqrt{\gamma\delta}}{e^{\frac{1}{2}(\mu+s\Phi^{-1}(y(1-T)+T))}}$
Translog $x(p) = \frac{\gamma - \eta \log p}{p}$	$Q(y) = \eta (\mathcal{W}(\underline{\varphi} e^{1+\frac{\gamma}{\eta}} (1-y)^{-\frac{1}{k}}) - 1)$	$Q(y) = \eta (\mathcal{W}(e^{1+\frac{\gamma}{\eta} + \mu + s\Phi^{-1}(y(1-T)+T)}) - 1)$

^a $\Phi(\cdot)$: c.d.f. of a standard normal; $\mathcal{W}(\cdot)$: the Lambert function.

TABLE B.V

QUANTILE FUNCTIONS FOR MARKUP DISTRIBUTIONS IMPLIED BY ASSUMPTIONS ABOUT PRODUCTIVITY (PARETO (P) OR TRUNCATED LOGNORMAL (tLN)) AND DEMAND (CREMR, LINEAR, LES, OR TRANSLOG)^a

Demand Function	Pareto Productivity $G_{\mathcal{P}}(\varphi) = 1 - \varphi^k \varphi^{-k}$	Truncated Lognormal Productivity $G_{tLN}(\varphi) = \frac{\Phi(\frac{\log \varphi - \mu}{\sigma}) - T}{1 - T}$
CREMR $p(x) = \frac{\beta}{x}(x - \gamma)^{\frac{\sigma-1}{\sigma}}$	$Q(y) = \frac{\sigma}{\sigma-1} \frac{1}{(1 + \omega^{-\sigma} (1-y)^{\frac{\sigma}{k}})}$ $\omega = \frac{\varphi \beta}{\gamma} \frac{\sigma-1}{\sigma}$	$Q(y) = \frac{\sigma}{\sigma-1} \frac{1}{1 + e^{-\tilde{\mu} - s\sigma\Phi^{-1}(y(1-T)+T)}}$ $\tilde{\mu} = \sigma(\mu - \log(\frac{\sigma}{\sigma-1} \frac{\gamma^{\frac{1}{\sigma}}}{\beta}))$
Linear $p(x) = \alpha - \beta x$	$Q(y) = \frac{1}{2}(1 + \omega(1-y)^{-\frac{1}{k}})$ $\omega = \alpha \varphi$	$Q(y) = \frac{1}{2}(1 + e^{\tilde{\mu} + s\Phi^{-1}(y(1-T)+T)})$ $\tilde{\mu} = \mu + \log \alpha$
LES $p(x) = \frac{\delta}{x+\gamma}$	$Q(y) = \sqrt{\omega(1-y)^{-1/k}}$ $\omega = \frac{\delta \varphi}{\gamma}$	$Q(y) = e^{\frac{1}{2}(\tilde{\mu} + s\Phi^{-1}(y(1-T)+T))}$ $\tilde{\mu} = \mu + \log(\frac{\delta}{\gamma})$
Translog $x(p) = \frac{\gamma - \eta \log p}{p}$	$Q(y) = \mathcal{W}(\omega(1-y)^{-1/k})$ $\omega = e^{1 + \frac{\gamma}{\eta}} \varphi$	$Q(y) = \mathcal{W}(e^{\tilde{\mu} + s\Phi^{-1}(y(1-T)+T)})$ $\tilde{\mu} = \mu + 1 + \frac{\gamma}{\eta}$

^a $\Phi(\cdot)$: c.d.f. of a standard normal; $\mathcal{W}(\cdot)$: the Lambert function.

TABLE B.VI

QQ ESTIMATOR FOR INDIAN SALES AND MARKUPS

	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

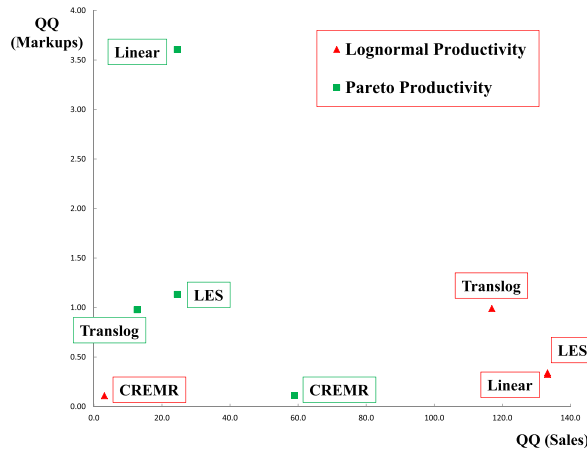


FIGURE B.4.—QQ estimator for Indian sales and markups.

TABLE B.VII

KLD FOR FRENCH EXPORTS COMPARED WITH PREDICTIONS FROM SELECTED DEMAND FUNCTIONS AND PRODUCTIVITY DISTRIBUTIONS

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

B.6. French Exports to Germany

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

As in Section B.3, we use the KLD as the criterion to determine how well different assumptions fit the data. Table B.VII gives the values of the KLD measuring the divergence from the empirical sales distribution from the distributions implied by CREMR/CES, translog, and linear demand functions combined with either Pareto or lognormal productivities. These distributions are calculated by combining the relevant productivity distribution with the relationships between productivity and sales given in Table I. (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. (The data are again normalized by the value of the KLD for a uniform distribution, which for this data set is 6.8082.)

Turning to the results in Table B.VII, the values of the minimized KLD show that, conditional on CREMR or CES demands, the lognormal provides a better overall fit than the Pareto: 0.0001 as opposed to 0.0012. However, the difference between distributions turns out to be much less significant than those between different specifications of demand. The KLD values for the translog and linear/LES specifications are much higher than for the CREMR case, as shown in the third and fourth columns of Table B.VII, 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 B.3, especially when we exclude the smallest firms as in Section B.4.

Table B.VIII repeats for French exports data the bootstrapping comparisons presented in Tables B.II and B.III for Indian sales and markup data, respectively. It is clear that the comparisons between different values of the KLD for the French data shown in Table B.VII are just as robust as those for the Indian data shown in Table B.I and Figure B.1 in the text.

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

TABLE B.VIII
BOOTSTRAPPED ROBUSTNESS OF THE KLD RANKING: FRENCH SALES^a

	CREMR + \mathcal{LN}	CREMR + \mathcal{P}	Translog + \mathcal{P}	Linear + \mathcal{P}	Translog + \mathcal{LN}	Linear + \mathcal{LN}
CREMR + \mathcal{LN}	–	0%	0%	0%	0%	0%
CREMR + \mathcal{P}	100%	–	0%	0%	0%	0%
Translog + \mathcal{P}	100%	100%	–	0%	0%	0%
Linear + \mathcal{P}	100%	100%	100%	–	0%	0%
Translog + \mathcal{LN}	100%	100%	100%	100%	–	0.3%
Linear + \mathcal{LN}	100%	100%	100%	100%	99.7%	–

^aSee text for explanation.

REFERENCES

- ARMENTER, R., AND M. KOREN (2014): “A Balls-and-Bins Model of Trade,” *American Economic Review*, 104, 2127–2151. [4]
- CHANEY, T. (2008): “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” *American Economic Review*, 98, 1707–1721. [3]
- CHRISTENSEN, L. R., D. W. JORGENSEN, AND L. J. LAU (1975): “Transcendental Logarithmic Utility Functions,” *American Economic Review*, 65, 367–383. [2]
- ELLISON, G., AND E. L. GLAESER (1997): “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach,” *Journal of Political Economy*, 105, 889–927. [4]
- HEAD, K., T. MAYER, AND M. THOENIG (2014): “Welfare and Trade Without Pareto,” *American Economic Review, Papers and Proceedings*, 104, 310–316. [7,8,10]
- KRATZ, M., AND S. I. RESNICK (1996): “The QQ-Estimator and Heavy Tails,” *Stochastic Models*, 12, 699–724. [7]
- KULLBACK, S., AND R. A. LEIBLER (1951): “On Information and Sufficiency,” *Annals of Mathematical Statistics*, 22, 79–86. [4]
- MRÁZOVÁ, M., AND J. P. NEARY (2017): “Not so Demanding: Demand Structure and Firm Behavior,” *American Economic Review*, 107, 3835–3874. [2-4]
- MUELLBAUER, J. (1975): “Aggregation, Income Distribution and Consumer Demand,” *Review of Economic Studies*, 42, 525–543. [2]
- NEVO, A. (2002): “Sample Selection and Information-Theoretic Alternatives to GMM,” *Journal of Econometrics*, 107, 149–157. [4]
- NIGAI, S. (2017): “A Tale of Two Tails: Productivity Distribution and the Gains From Trade,” *Journal of International Economics*, 104, 44–62. [8]

Co-editor Aviv Nevo handled this manuscript.

Manuscript received 19 June, 2019; final version accepted 19 November, 2020; available online 3 December, 2020.