On the Modeling of Size Distributions When Technologies Are Complex

Jakub Growiec*

April 14, 2013

Preliminary and incomplete.

Abstract. The current paper considers a stochastic R&D process where the invented production technologies consist of a large number n of complementary components. Complementarity is modeled with the CES aggregator function. Drawing from the Central Limit Theorem and the Extreme Value Theory we find, under very general assumptions, that the cross-sectional distributions of technological productivity are well-approximated either by the log-normal, Weibull, or a novel "CES/Normal" distribution, depending on the underlying elasticity of substitution between technology components. We numerically assess the rate of convergence of the true unit factor productivity distribution to the theoretical limit with n.

Keywords: distribution of technological productivity, stochastic R&D, CES, Weibull distribution, log-normal distribution, limiting distribution

JEL Classification Numbers: E23, L11, O47

*Warsaw School of Economics, Institute of Econometrics, al. Niepodległości 162, 02-554 Warszawa, Poland, and National Bank of Poland, Economic Institute. E-mail: jakub.growiec@sgh.waw.pl.

1 Introduction

Most technologies used nowadays are complex in the sense that the production processes (and products themselves) consist of a large number of components which might interact with each other in complementary ways (e.g., Kremer, 1993; Blanchard and Kremer, 1997; Jones, 2011). Based on this insight, the current paper assumes that the total productivity of any given technology is functionally dependent on the individual productivities of its n components as well as the elasticity of substitution between them, σ. This functional relationship is captured by the CES aggregator function. The stochastic R&D process which invents new complex technologies is in turn assumed to consist in drawing productivities of the components from certain predefined probability distributions (Jones, 2005; Growiec, 2008a,b, 2012).

Based on this set of assumptions, we obtain surprisingly general results regarding the implied cross-sectional distributions of technological productivity. Namely, drawing from the Central Limit Theorem and the Extreme Value Theory, we find that if the number of components of a technology, n, is sufficiently large, these distributions should be well approximated either by:

(i) the log-normal distribution - in the case of unitary elasticity of substitution between the components ($\sigma = 1$);

(ii) the Weibull distribution - in the case of perfect complementarity between the components (the "weakest link" assumption, $\sigma = 0$),

(iii) the Gaussian distribution - in the (empirically very unlikely) case of perfect substitutability between the components ($\sigma \to \infty$),

(iv) a novel "CES/Normal" distribution - in any intermediate CES case, parametrized by the elasticity of substitution between the components ($\sigma > 0, \sigma \neq 1$).

Our theoretical contribution to the literature is supplemented with a series of numerical simulations, allowing us to approximate the rate of convergence of the true distribution to the theoretical limit with n. We also numerically assess the dependence of the limiting "CES/Normal" distribution on the degree of complementarity between the technology components, σ.

Potential empirical applications of the theoretical result, reaching beyond the scope of the current paper, include providing answers to the following research questions:

- Does the "CES/Normal" distribution derived here (eq. (10)) fit the data on firm sizes, sales, R&D spending, etc.? What is the implied value of σ?

- Do industries differ in terms of their technology complexity as captured by n?
- Do industries differ in terms of the complementarity of technology components as captured by σ?
- How do firms' optimal technology choices and production function aggregation (Growiec, 2012) enter the picture?

The remainder of the paper is structured as follows. Section 2 sets up and discusses the model. Section 3 presents the numerical results. Section 4 concludes.

2 The model

2.1 Distributions of complex technologies

The point of departure of the current model is the assumption that technologies, invented within the R&D process, are inherently complex and consist of a large number of complementary components. Formally, this can be written down in the following way.

Assumption 1 *The R&D process determines the productivity of any newly invented technology Y as a constant elasticity of substitution (CES) aggregate over $n \in \mathbb{N}$ independent draws X_i, $i = 1, ..., n$, from the elementary idea distribution $\mathcal{F}$:*

$$Y = \begin{cases} \min\{X_i\}_{i=1}^n, & \theta = -\infty, \\ \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^{1/\theta}, & \theta \in (-\infty, 0) \cup (0, 1], \\ \prod_{i=1}^n X_i^{1/n}, & \theta = 0. \end{cases} \tag{1}$$

The elementary distribution $\mathcal{F}$ is assumed to have a positive density on $[w, v]$ and zero density otherwise (where $w \geq 0$ and $v > w$ can be infinite), and satisfy the condition of a regularly varying lower tail (Leadbetter et al., 1983):

$$\lim_{p\to 0_+} \frac{\mathcal{F}(w + px)}{\mathcal{F}(w + p)} = x^\alpha \tag{2}$$

for all $x > 0$ and a certain $\alpha > 0$.

The parameter n in the above assumption captures the number of constituent components of any given (composite) technology, and thus measures the complexity of any state-of-the-art technology. The substitutability parameter θ is related to the elasticity of substitution σ via $\theta = \frac{\sigma-1}{\sigma}$, or $\sigma = \frac{1}{1-\theta}$. The case $\theta < 0$ captures the case where the components of technologies are gross complements ($\sigma \in [0,1)$), whereas $\theta \in (0,1]$ implies that they are gross substitutes ($\sigma > 1$).

It should be noted at this point that, as argued repeatedly by Kremer (1993), Jones (2011) and Growiec (2012), the gross complementarity case is much more likely to provide an adequate description of real-world production processes than the gross subsitutability case. The example of the explosion of the space shuttle *Challenger* due to a failure of an inexpensive O-ring, put forward by Kremer (1993), is perhaps the best possible illustration of the potentially complementary character of components of complex technologies.

More precisely, the *minimum* case (a Leontief function) reflects the extreme case where technology components are *perfectly* complementary, and thus the actual productivity of a complex idea is determined by the productivity of its "weakest link" (or "bottleneck"). This case was assumed in the earlier related contribution by Growiec (2012). Although likely, this case need not hold exactly in reality, since certain deficiencies of design can often be covered by advantages in different respects. The more general CES case captures exactly this possibility (see also Klump et al., 2012).

The limiting Cobb-Douglas case ($\theta = 0$) is the threshold case delineating gross complementarity from gross substitutability. As shown by Kremer (1993), this case is already quite illustrative of effects of complementarity between components of technologies.

Although technical in nature, the assumption (2) imposed on elementary probability distributions $\mathcal{F}$ can also be intepreted in economic terms. First, the support of the distribution must be bounded from below, which means researchers are not allowed to draw infinitely "bad" technologies (zero is a natural lower bound). This rules out distributions defined on the whole $\mathbb{R}$ such as the Gaussian. Second, the pdf of the distribution $\mathcal{F}$ cannot increase smoothly from zero at w; there must be a jump. This means that the probability of getting a draw which is "as bad as it gets" cannot be negligible, and this rules out a few more candidate distributions such as the lognormal or the Fréchet. Third, the lowest possible value of the random variable cannot be an isolated atom, which rules out all discrete distributions such as the two-point distribution, the binomial, negative binomial, Poisson, etc. Yet, the set of distributions satisfying (2) is still reasonably large. It includes, among others, the frequently assumed Pareto, uniform, truncated Gaussian, and Weibull distributions (cf. Growiec, 2012).

2.2 Theoretical results

Let us now derive our theoretical results related to the limiting distribution of Y when the number of technology components $n \to \infty$. For analytical convenience, we denote $\mu_\theta = EX_i^\theta$ and $\sigma_\theta = \sqrt{D^2(X_i^\theta)} = \sqrt{EX_i^{2\theta} - (EX_i^\theta)^2}$.

Letting the technology complexity n be arbitrarily large, we obtain the following result:

Proposition 1 *If Assumption 1 holds with $\theta = -\infty$ ($\sigma = 0$), then as $n \to \infty$, the minimum of n independent random draws from the distribution $\mathcal{F}$, after appropriate normalization, converges in distribution to the Weibull distribution with the shape parameter α:*

$$[1 - \mathcal{F}(xp_n + w)]^n \xrightarrow{d} e^{-\left(\frac{x}{\lambda}\right)^\alpha}, \tag{3}$$

where $w = \inf\{x \in \mathbb{R} : \mathcal{F}(x) > 0\}$, $p_n = \frac{1}{\lambda}\left(\mathcal{F}^{-1}\left(\frac{1}{n}\right) - w\right)$ and the free parameter $\lambda > 0$ is assumed to be proportional to the mean of the underlying distribution $\mathcal{F}$.

Proposition 2 *If Assumption 1 holds with $\theta = 0$ ($\sigma = 1$), then as $n \to \infty$, the product of n independent random draws from the distribution $\mathcal{F}$, after appropriate normalization, converges in distribution to the log-normal distribution:*

$$[1 - \mathcal{F}(x)]^n \xrightarrow{d} 1 - \Phi\left(\frac{\ln x - \mu_1}{\sigma_1}\right). \tag{4}$$

Proposition 3 *If Assumption 1 holds with $\theta \in (-\infty, 0) \cup (0, 1]$ ($\sigma \in (0, 1) \cup (1, +\infty]$), then as $n \to \infty$, the CES bundle of n independent random draws from the distribution $\mathcal{F}$, after appropriate normalization, converges in distribution to the "CES/Normal" distribution with complementary cdf:*

$$[1 - \mathcal{F}(x)]^n \xrightarrow{d} \Phi\left(\frac{x^\theta - \mu_\theta}{\sigma_\theta}\right), \tag{5}$$

and thus the following pdf:

$$g(x) = \frac{|\theta|}{\sigma_\theta\sqrt{2\pi}} x^{\theta-1} e^{-\frac{(x^\theta - \mu_\theta)^2}{2\sigma_\theta^2}}, \qquad x > 0. \tag{6}$$

Hence, the class of "CES/Normal" distributions encompasses the Gaussian distribution for the limiting case $\theta = 1$ where the technology components are perfectly substitutable.

Proof of Propositions 1–3. For the case $\theta = -\infty$ the proposition follows directly from the Fisher–Tippett–Gnedenko extreme value theorem, applied to the distribution

$\mathcal{F}$ (Theorem 1.1.3 in de Haan and Ferreira, 2006, rephrased so that it captures the minimum instead of maximum). From the theorem specifying the domain of attraction of the Weibull distribution (Theorem 1.2.1 in de Haan and Ferreira, 2006; Section 1.3 in Kotz and Nadarajah, 2000), we obtain the necessary and sufficient conditions for the complementarity mechanism to work. The implied parameter α is found to be unitary for a wide range of distributions $\mathcal{F}$ (Growiec, 2012).

With $\theta = 0$, by the Central Limit Theorem the distribution of Y satisfies:

$$\left(\frac{\prod_{i=1}^{n} X_i}{e^{n\mu_1}}\right)^{\frac{1}{\sqrt{n}}} \xrightarrow{d} \log N(0, \sigma_1^2), \tag{7}$$

where $\mu_1 = E(\ln X_i)$ and $\sigma_1 = E(\ln X_i)^2 - (E \ln X_i)^2$.

For the case $\theta \in (-\infty, 0) \cup (0, 1]$, we may use the Central Limit Theorem again, obtaining:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{\infty} X_i^{\theta} - \mu_\theta\right) \to N(0, \sigma_\theta^2). \tag{8}$$

Thus, for sufficiently large n we may write that $X(n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\infty} X_i^{\theta} \sim N(\sqrt{n}\mu_\theta, \sigma_\theta^2)$, with slight abuse of notation. The mean of this distribution, $\sqrt{n}\mu_\theta$, is then also sufficiently large for the probability that $X(n) < 0$ to be negligible. The CLT approximation is thus consistent with the underlying assumption that $X_i \sim \mathcal{F}$ takes only positive values. It therefore makes sense to calculate the limiting distribution of $Y = \lim_{n\to\infty} X(n)^{1/\theta}$.

This limiting distribution takes the following form. For $y > 0$ and with $\theta < 0$, the cdf of Y satisfies:

$$\mathcal{G}(y) = P(Y \le y) = \lim_{n\to\infty} P(X(n) \ge y^{\theta}) = 1 - \bar{\Phi}\left(\frac{y^{\theta} - \mu_\theta}{\sigma_\theta}\right), \tag{9}$$

where the cdf of the truncated normal distribution $\bar{\Phi}\left(\frac{x-\mu_\theta}{\sigma_\theta}\right) = \frac{\Phi\left(\frac{x-\mu_\theta}{\sigma_\theta}\right) - \Phi\left(\frac{-\mu_\theta}{\sigma_\theta}\right)}{1-\Phi\left(\frac{-\mu_\theta}{\sigma_\theta}\right)} \approx \Phi\left(\frac{x-\mu_\theta}{\sigma_\theta}\right)$ if n is sufficiently large. Upon differentiation, we obtain the following pdf of the limiting "CES/Normal" distribution, parametrized by $\theta < 0, \mu_\theta > 0$ and $\sigma_\theta > 0$:

$$g(y) = \frac{-\theta}{\sigma_\theta\sqrt{2\pi}} \frac{y^{\theta-1}}{1 - \Phi\left(\frac{-\mu_\theta}{\sigma_\theta}\right)} e^{-\frac{(y^{\theta}-\mu_\theta)^2}{2\sigma_\theta^2}} \approx \frac{-\theta}{\sigma_\theta\sqrt{2\pi}} y^{\theta-1} e^{-\frac{(y^{\theta}-\mu_\theta)^2}{2\sigma_\theta^2}}, \qquad y > 0. \tag{10}$$

Conversely, for $y > 0$ and with $\theta \in (0, 1]$, the cdf of Y satisfies:

$$\mathcal{G}(y) = P(Y \le y) = \lim_{n\to\infty} P(X(n) \le y^{\theta}) = \bar{\Phi}\left(\frac{y^{\theta} - \mu_\theta}{\sigma_\theta}\right) \approx \Phi\left(\frac{y^{\theta} - \mu_\theta}{\sigma_\theta}\right). \tag{11}$$

Upon differentiation, we obtain the following pdf of the limiting "CES/Normal" distribution, parametrized by $\theta \in (0,1], \mu_\theta > 0$ and $\sigma_\theta > 0$:

$$g(y) = \frac{\theta}{\sigma_\theta \sqrt{2\pi}} \frac{y^{\theta-1}}{1-\Phi\left(\frac{-\mu_\theta}{\sigma_\theta}\right)} e^{-\frac{(y^\theta-\mu_\theta)^2}{2\sigma_\theta^2}} \approx \frac{\theta}{\sigma_\theta\sqrt{2\pi}} y^{\theta-1} e^{-\frac{(y^\theta-\mu_\theta)^2}{2\sigma_\theta^2}}, \qquad y > 0.\blacksquare \tag{12}$$

3 Numerical results

The upside of above theoretical results is that they provide theoretical limits for the distributions of complex technologies, regardless of the underlying distribution of technology components $\mathcal{F}$. Unfortunately, these limits are exactly correct only if the technologies are *infinitely complex*. It is therefore of great importance to assess the pace of convergence of true underlying distributions to the limiting ones with the number of components, n, to see how large departures from the theoretical limit should be expected if n is in fact finite. To this end, we have carried out a series of numerical computations.

Another interesting issue is the dependence of the limiting "CES/Normal" distribution on the complementarity parameter θ (or equivalently, the elasticity of substitution, σ). This will be assessed numerically as well.

In the current section, we shall first describe our numerical framework and then pass on to a short outline of the results.

3.1 Generating the distribution of Y

The preliminary step of our numerical exercises consists in generating a sample of n units, randomly and independently drawn from the *uniform* distribution defined on an interval in the positive domain (which is a particular instance of a distribution satisfying Assumption 1):

$$X_i \sim U[a,b], \qquad b > a > 0, \; i = 1, 2, ..., n. \tag{13}$$

Next, we computed the CES aggregate of these random draws according to:

$$Y = \begin{cases} \min_{i=1}^n \{X_i\}, & \theta = -\infty, \\ \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^{1/\theta}, & \theta \in (-\infty, 0) \cup (0,1], \\ \prod_{i=1}^n X_i^{1/n}, & \theta = 0, \end{cases} \tag{14}$$

for a few arbitrary values of the complementarity parameter θ. For a fixed sample size of n, we repeated this procedure $m = 10000$ times and plotted the empirical histogram of Y.

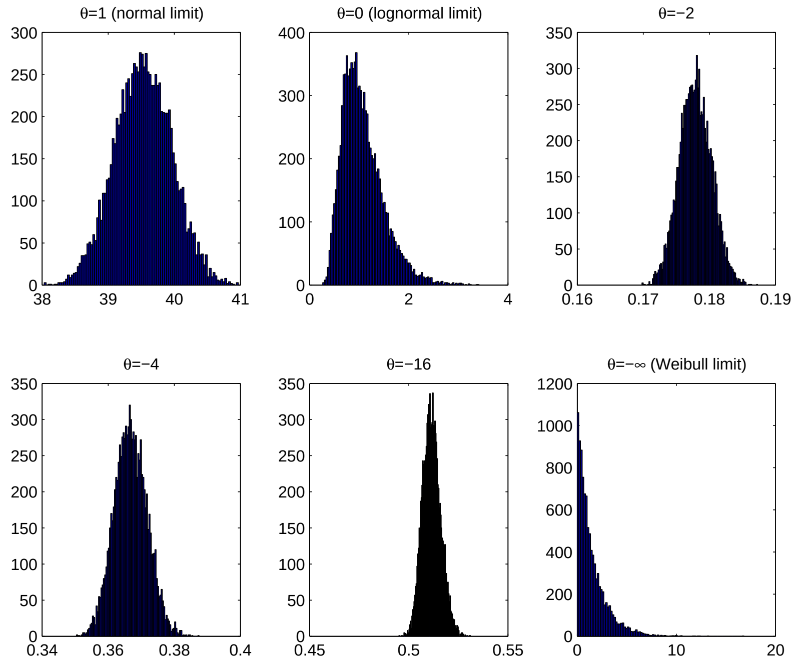

Figure 1: Histograms of simulated Y. Assumed parameter values: $n = 1000, a = 0.5, b = 2$.

The histograms of the generated variables are presented in Figure 1. These histograms (with $B = 100$ bins) were then transformed into empirical pdfs.

Figure 1 confirms a trend of increasing skewness of the empirical distribution when θ declines towards $-\infty$, which can be also inferred from the pdf of the limiting "CES/Normal" distribution.

3.2 The lognormal and Weibull limits

The second step of our numerical exercise consists in confirming the theoretical lognormal limit for the case $\theta = 0$. We see that the theoretical distribution indeed aligns with the simulated data almost perfectly when $n = 1000$. Similarly, the theoretical Weibull limit obtained for the case $\theta = -\infty$ fits the simulated data almost perfectly as well.[1] The results are contained in Figures 2–3.

[1] Nonlinear least squares fit of the Generalized Gamma distribution (a three-parameter class of distributions) is included for comparison.

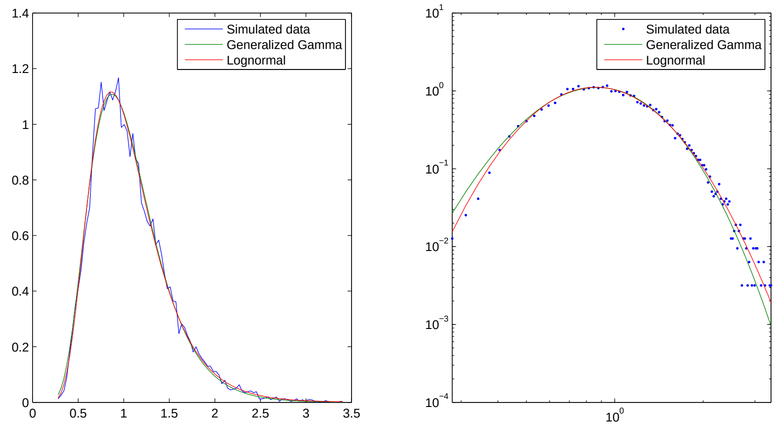

Figure 2: Simulated pdf of Y vs. the lognormal limit for $n \to \infty$. Assumed parameter values: $n = 1000, a = 0.5, b = 2$.

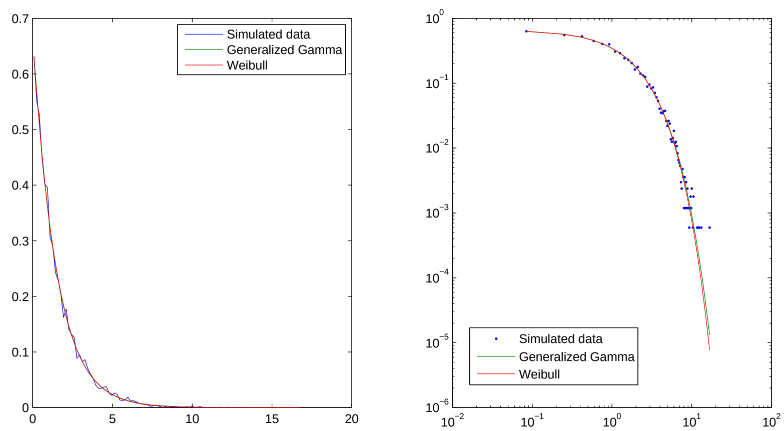

Figure 3: Simulated pdf of Y vs. the Weibull limit for $n \to \infty$. Assumed parameter values: $n = 1000, a = 0.5, b = 2$.

3.3 The CES/Normal limit

The next step consisted in confirming the theoretical formula for the limiting "CES/Normal" distribution under intermediate values of the complementarity parameter, $\theta \in (-\infty, 0) \cup (0, 1]$. As it can be seen on Figure 4, the theoretical "CES/Normal" limit is fully confirmed in this case; it is clear that none of the seemingly similar (and more generously parametrized) distributions can be fitted to the simulated data equally well.

This numerical exercise confirms that (a) the Weibull distribution misses the shape of the pdf completely when θ is finite, (b) all other considered distributions fit these data

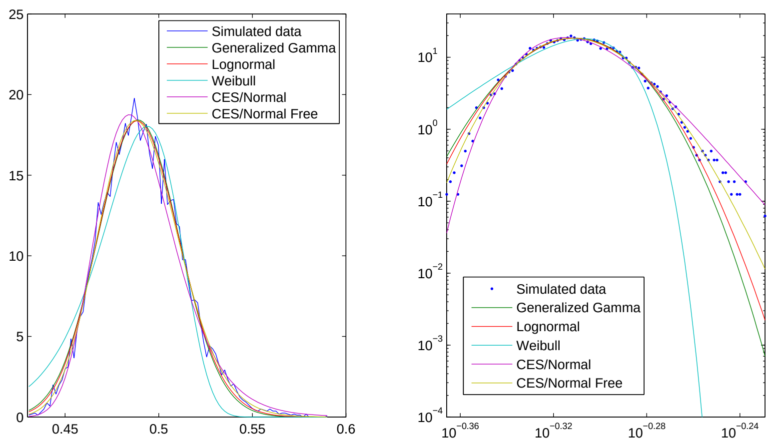

Figure 4: Simulated pdf of Y vs. the CES/Normal limit (eq. (10)) for $n \to \infty$. Assumed parameter values: $n = 100, a = 0.5, b = 2, \theta = -4$.

better, but tend to underestimate the probabilities of tail events. In fact, the actual limiting distribution is much more skewed than the estimated pdfs.

Regarding Figure 4, please note the following difference between the "CES/Normal" distribution and the "CES/Normal Free" case. The first one takes the (known) theoretical values of mean and variance $(\mu_\theta, \sigma_\theta)$ as well as θ itself as given, whereas the latter takes them as free parameters to be estimated by nonlinear least squares. We see that improving the fit in the body of the distribution of the finite CES aggregate ($n = 100 < \infty$) compromises the quality of fit in the right tail.

3.4 Convergence as $n \to \infty$

The next step consists in repeating the numerical experiment for a fixed value of $\theta \in (-\infty, 0) \cup (0, 1]$ but various sample sizes n to assess the pace of convergence of the resultant distribution to the theoretical "CES/Normal" limiting distribution (eq. (10)). This requires us to standardize the resultant distributions so that they have a fixed (e.g., zero) mean and unitary standard deviation.

In Figure 5 we observe that as n increases, the resulting distribution gradually evolves from the uniform distribution of X_i to the limiting CES/Normal distribution, derived in the previous section. In the body of the distribution, convergence is rather fast and is largely done already for $n = 16$. The tails of the distribution are however much thinner for small n than in the limiting distribution. This mirrors the known fact that tails of a distribution need much more time to take their final shape, because they are by definition capturing rare events. Even for $n = 2000$, although the fit in the body is already perfect, no observations have been found for tail events exceeding 6.

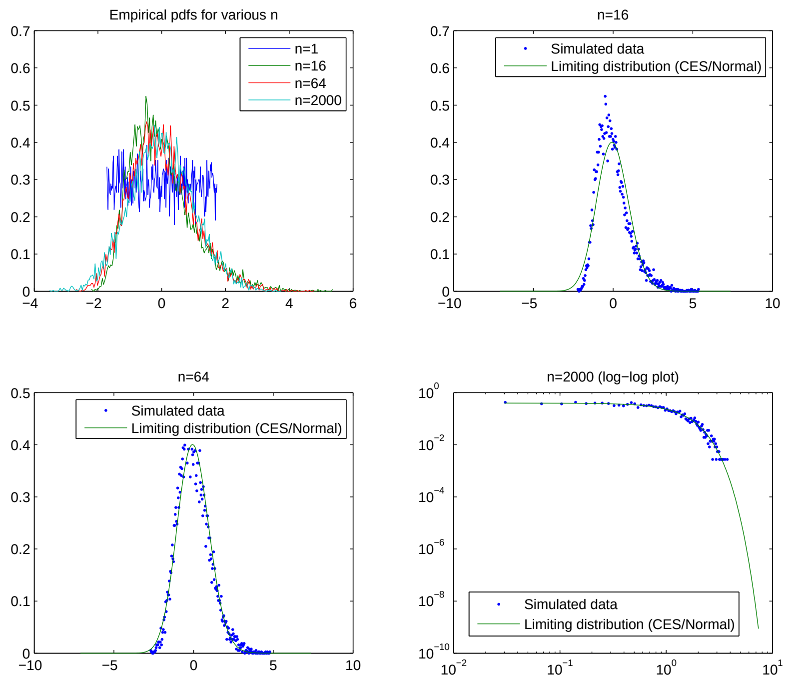

Figure 5: Convergence of Y to the CES/Normal limit (eq. (10)) for different values of n. Assumed parameter values: $a = 0.5, b = 2, \theta = -4$.

3.5 Dependence of the CES/Normal distribution on θ

The final step of the numerical exercise is to illustrate the dependence of the limiting CES/Normal distribution on the complementarity parameter θ, with special reference to the tail. As illustrated by Figure 6, we find that the larger is complementarity between technology components, the fatter the tail of the limiting distribution. To obtain this result, we again standardize the resultant distributions.

Please note that the tail of the distribution with greatest complementarity is probably misspecified due to numerical error.

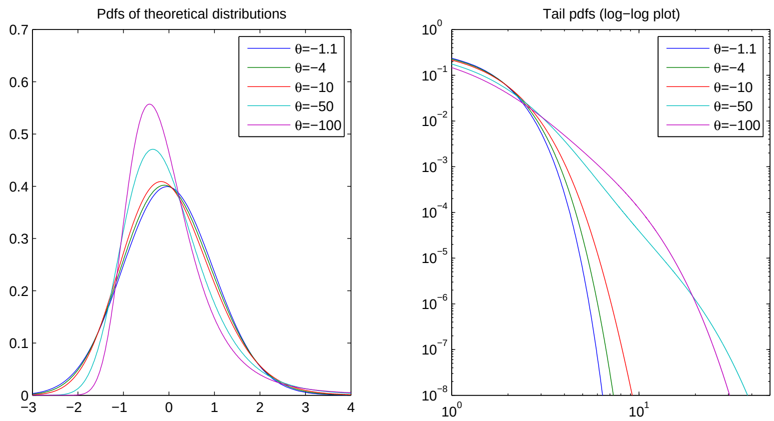

Figure 6: Convergence of Y to the CES/Normal limit (eq. (10)) for different values of θ. Assumed parameter values: $a = 0.5, b = 2$.

4 Conclusion

In the current paper, we have obtained surprisingly general results regarding the implied cross-sectional distributions of technological productivity. Namely, drawing from the Central Limit Theorem and the Extreme Value Theory, we find that if the number of components of a technology, n, is sufficiently large, these distributions should be well approximated either by:

(i) the log-normal distribution - in the case of unitary elasticity of substitution between the components ($\sigma = 1$);

(ii) the Weibull distribution - in the case of perfect complementarity between the components (the "weakest link" assumption, $\sigma = 0$),

(iii) the Gaussian distribution - in the (empirically very unlikely) case of perfect substitutability between the components ($\sigma \to \infty$),

(iv) a novel "CES/Normal" distribution - in any intermediate CES case, parametrized by the elasticity of substitution between the components ($\sigma > 0, \sigma \neq 1$).

Our theoretical contribution to the literature has been supplemented with a series of numerical simulations, allowing us to approximate the rate of convergence of the true distribution to the theoretical limit with n. We have also numerically assessed the dependence of the limiting "CES/Normal" distribution on the degree of complementarity between the technology components, σ.

What remains on the research agenda is to:

- consider the possibility of using normalized CES aggregator functions (cf. Klump et al., 2012) – or other reparametrizations – instead of standard CES ones in computations of the "CES/Normal" limit,
- discuss the implied moments of the limiting distributions; ensure that there is convergence in distribution when $\theta \to 0$ or $\theta \to -\infty$,
- provide approximate theoretical results on the pace of convergence of n-unit technologies to the "CES/Normal", Weibull or log-normal limit,
- verify the empirical relevance of CES/Normal distributions. Do we find it in data on firm sizes, sales, R&D spending, etc.? What are the implied values of σ?

References

[1] Blanchard, O., M. Kremer (1997), "Disorganization," *Quarterly Journal of Economics* 112(4), 1091-1126.

[2] de Haan, L., A. Ferreira (2006), *Extreme Value Theory. An Introduction*. Springer.

[3] Growiec, J. (2008a), "A New Class of Production Functions and an Argument Against Purely Labor-Augmenting Technical Change", *International Journal of Economic Theory* 4(4), 483-502.

[4] Growiec, J. (2008b), "Production Functions and Distributions of Unit Factor Productivities: Uncovering the Link", *Economics Letters* 101(1), 87-90.

[5] Growiec, J. (2012), "A Microfoundation for Normalized CES Production Functions with Factor-Augmenting Technical Change", NBP Working Paper 98.

[6] Jones, C.I. (2005), "The Shape of Aggregate Production Functions and the Direction of Technical Change", *Quarterly Journal of Economics* 120(2), 517-549.

[7] Jones, C.I. (2011), "Intermediate Goods and Weak Links in the Theory of Economic Development", *American Economic Journal: Macroeconomics* 3(2), 1-28.

[8] Klump, R., P. McAdam, A. Willman (2012), "The Normalized CES Production Function: Theory and Empirics", *Journal of Economic Surveys* 26(5), 769-799.

[9] Kotz, S., S. Nadarajah (2000), *Extreme Value Distributions. Theory and Applications*. Imperial College Press, London.

[10] Kremer, M. (1993), "The O-Ring Theory of Economic Development", *Quarterly Journal of Economics* 108(3), 551-575.

[11] Leadbetter, M. R., G. Lindgren, H. Rootzén (1983), *Extremes and Related Properties of Random Sequences and Processes*. Springer.