The Shape of Production Functions and the Direction of Technical Change

Charles I. Jones*

Department of Economics, U.C. Berkeley and NBER
E-mail: chad@econ.berkeley.edu
http://elsa.berkeley.edu/˜chad

November 5, 2004 – Version 2.0

This paper views the standard production function in macroeconomics as a reduced form and derives its properties from microfoundations. The shape of this production function is governed by the distribution of ideas. If that distribution is Pareto, then two results obtain: the global production function is Cobb-Douglas, and technical change in the long run is labor-augmenting. Kortum (1997) showed that Pareto distributions are necessary if search-based idea models are to exhibit steady-state growth. Here we show that this same assumption delivers the additional results about the shape of the production function and the direction of technical change.

Key Words: Cobb-Douglas, Pareto, Power Law, Labor-Augmenting Technical Change, Steady State Growth
JEL Classification: O40, E10

*I am grateful to Daron Acemoglu, Susanto Basu, Francesco Caselli, Hal Cole, Xavier Gabaix, Doug Gollin, Pete Klenow, Mike Scherer, Robert Solow, Alwyn Young, and seminar participants at Arizona State University, the Chicago GSB, Harvard, the San Francisco Fed, Stanford, U.C. Berkeley, UCLA, U.C. Davis, and U.C. Santa Cruz for comments. Sam Kortum provided especially useful insights, for which I am most appreciative. Meredith Beechey, Robert Johnson, and Dean Scrimgeour supplied excellent research assistance. This research is supported by NSF grant SES-0242000.

1. INTRODUCTION

Much of macroeconomics — and an even larger fraction of the growth literature — makes strong assumptions about the shape of the production function and the direction of technical change. In particular, it is well-known that for a neoclassical growth model to exhibit steady-state growth, either the production function must be Cobb-Douglas or technical change must be labor augmenting in the long run. But apart from analytic convenience, is there any justification for these assumptions?

Where do production functions come from? To take a common example, our models frequently specify a relation $y = f(k, \cdot)$ that determines how much output per worker y can be produced with any quantity of capital per worker k. We typically assume the economy is endowed with this function, but consider how we might derive it from deeper microfoundations.

Suppose production techniques are ideas that get discovered over time. One example of such an idea would be a Leontief technology that says, "for each unit of labor, take k^* units of capital. Follow these instructions [omitted], and you will get out y^* units of output." The values k^* and y^* are parameters of this production technique.

If one wants to produce with a very different capital-labor ratio from k^*, this Leontief technique is not particularly helpful, and one needs to discover a new idea "appropriate" to the higher capital-labor ratio.[1] Notice that one can replace the Leontief structure with a production technology that exhibits a low elasticity of substitution, and this statement remains true: to take advantage of a substantially higher capital-labor ratio, one really needs a new technique targetted at that capital-labor ratio. One needs a new idea.

According to this view, the standard production function that we write down, mapping the entire range of capital-labor ratios into output per

[1]This use of appropriate technologies is related to Atkinson and Stiglitz (1969) and Basu and Weil (1998).

worker, is a reduced form. It is not a single technology, but rather represents the substitution possibilities across different production techniques. The elasticity of substitution for this global production function depends on the extent to which new techniques that are appropriate at higher capital-labor ratios have been discovered. That is, it depends on the distribution of ideas.

But from what distribution are ideas drawn? Kortum (1997) examined a search model of growth in which ideas are productivity levels that are drawn from a distribution. He showed that the only way to get exponential growth in such a model is if ideas are drawn from a Pareto distribution, at least in the upper tail.

This same basic assumption, that ideas are drawn from a Pareto distribution, yields two additional results in the framework considered here. First, the global production function is Cobb-Douglas. Second, the optimal choice of the individual production techniques leads technological change to be purely labor-augmenting in the long run. In other words, an assumption Kortum (1997) suggests we make if we want a model to exhibit steady-state growth leads to important predictions about the shape of production functions and the direction of technical change.

In addition to Kortum (1997), this paper is most closely related to an older paper by Houthakker (1955-56) and to two recent papers, Acemoglu (2003b) and Caselli and Coleman (2004). The way in which these papers fit together will be discussed below.[2]

Section 2 of this paper presents a simple baseline model that illustrates all of the main results of this paper. In particular, that section shows how a specific shape for the technology menu produces a Cobb-Douglas produc-

[2]The insight that production techniques underlie what I call the global production function is present in the old reswitching debate; see Robinson (1953–1954). The notion that distributions for individual parameters aggregate up to yield a well-behaved function is also found in the theory of aggregate demand; see Hildenbrand (1983) and Grandmont (1987).

tion function and labor-augmenting technical change. Section 3 develops the full model with richer microfoundations and derives the Cobb-Douglas result, while Section 4 discusses the underlying assumptions and the relationship between this model and Houthakker (1955–1956). Section 5 develops the implications for the direction of technical change. Section 6 provides a numerical example of the model, and Section 7 concludes.

2. A BASELINE MODEL

2.1. Preliminaries

Let a particular production technique — call it technique i — be defined by two parameters, a_i and b_i. With this technique, output Y can be produced with capital K and labor L according to the *local production function* associated with technique i:

$$Y = \tilde{F}(b_i K, a_i L). \tag{1}$$

We assume that $\tilde{F}(\cdot,\cdot)$ exhibits an elasticity of substitution less than one between its inputs and constant returns to scale in K and L. In addition, we make the usual neoclassical assumption that $\tilde{F}$ possesses positive but diminishing marginal products and satisfies the Inada conditions.

This production function can be rearranged to give

$$Y = a_i L \tilde{F}\left(\frac{b_i K}{a_i L}, 1\right), \tag{2}$$

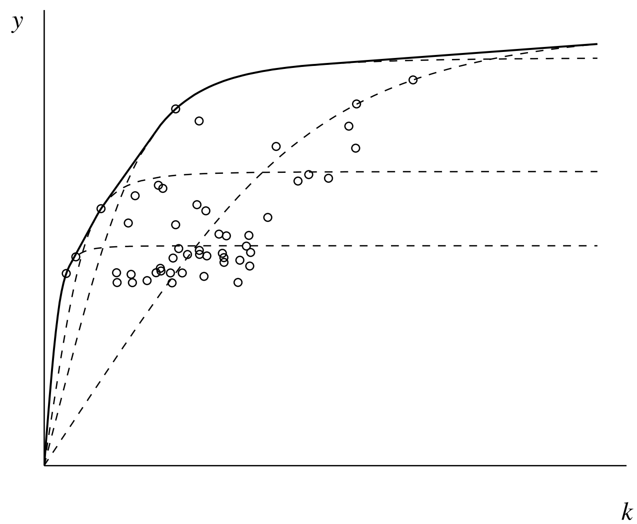
so that in per worker terms we have

$$y = a_i \tilde{F}\left(\frac{b_i}{a_i} k, 1\right), \tag{3}$$

where $y \equiv Y/L$ and $k \equiv K/L$. Now, define $y_i \equiv a_i$ and $k_i \equiv a_i/b_i$. Then the production technique can be written as

$$y = y_i \tilde{F}\left(\frac{k}{k_i}, 1\right). \tag{4}$$

FIGURE 1. An Example of the Global Production Function



Circles identify distinct production techniques; for some of these, the local production function associated with the technique has been drawn as a dashed line. The heavy solid line shows the convex hull of the local production functions, i.e. the global production function.

If we choose our units so that $\tilde{F}(1, 1) = 1$, then we have the nice property that $k = k_i$ implies that $y = y_i$. Therefore, we can think of technique i as being indexed by a_i and b_i, or, equivalently, by k_i and y_i.

The shape of the global production function is driven by the distribution of alternative production techniques rather than by the shape of the local production function that applies for a single technique.[3] To illustrate this, consider the example given in Figure 1. The circles in this figure denote different production techniques that are available — the set of (k_i, y_i) pairs. For a subset of these, we also plot the local production function

[3]Other models in the literature feature a difference between the short-run and long-run elasticities of substitution, as opposed to the local-global distinction made here. These include the putty-clay models of Caballero and Hammour (1998) and Gilchrist and Williams (2000).

$y = \tilde{F}(b_i k, a_i)$. Finally, the heavy solid line shows the global production function, given by the convex hull of the local production techniques. For any given level of k, the global production function shows the maximum amount of output per worker that can be produced using the set of ideas that are available.

The key question we'd like to answer is this: What is the shape of the global production function? To make progress, we now turn to a simple baseline model.

2.2. The Baseline Model

We begin with a simple model, really not much more than an example. However this baseline model turns out to be very useful: it is easy to analyze and captures the essence of the model with more detailed microfoundations that is presented in Section 3.

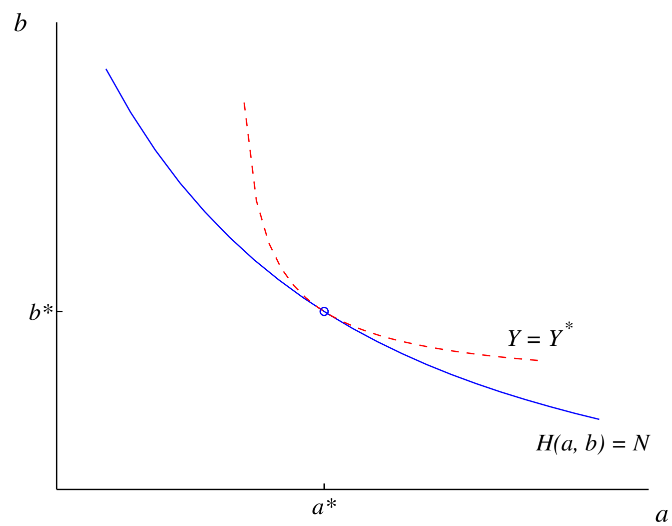
At any given point in time, a firm has a stock of ideas — a collection of local production techniques — from which to choose. This set of production techniques is characterized by the following technology menu:

$$H(a, b) = N \tag{5}$$

where $H_a > 0$, $H_b > 0$, and $N > 0$. Along this menu, there is a tradeoff: ideas with a high value of b are associated with a low value of a. N parameterizes the location of this technology menu and might be thought of as the level of knowledge. A higher N means the technology menu supports higher levels of a and b. Associated with any (a, b) pair from this technology menu is a local production function $Y = \tilde{F}(bK, aL)$, with the properties assumed above in equation (1), including an elasticity of substitution less than one and constant returns to scale in K and L.

The *global production function* for this firm describes the maximum amount of output the firm can produce from a particular set of inputs, when

FIGURE 2. Direction of Technical Change



it is free to choose any production technique from the technology menu. That is, the global production function $F(K, L; N)$ is defined as

$$Y = F(K, L; N) \equiv \max_{b,a} \ \tilde{F}(bK, aL) \tag{6}$$

subject to (5).

Characterizing the global production function is straightforward. Graphically, one version of this problem with an interior solution is shown in Figure 2. Algebraically, an interior solution equates the marginal rate of technical substitution along the isoquant to the marginal rate of technical substitution along the technology menu. We can express this in its elasticity form and use the fact that the elasticity of production with respect to b is the same as the elasticity with respect to K to get the following result:

$$\frac{\theta_K}{\theta_L} = \frac{\eta_b}{\eta_a}, \tag{7}$$

where $\theta_K(a, b; K, L) \equiv \tilde{F}_1 bK/Y$ is the capital share, $\theta_L = 1 - \theta_K$ is the analogous labor share, $\eta_b \equiv \frac{\partial H}{\partial b}\frac{b}{H}$ is the elasticity of H with respect to b, and η_a is the analogous elasticity with respect to a. The optimal technology choice equates the ratio of the capital and labor shares to the ratio of the elasticities of the technology menu.

In Figure 2, we drew the technology menu as convex to the origin. Of course, we could have drawn the curve as concave or linear, or we could have drawn it as convex, but with a sharper curvature than the isoquant. However, it turns out that the constant elasticity version of the convex curve delivers a particularly nice result.[4] In particular, suppose the technology menu is given by

$$H(a,b) \equiv a^{\alpha} b^{\beta} = N, \quad \alpha > 0,\ \beta > 0. \tag{8}$$

In this case, the elasticity $\eta_b/\eta_a = \beta/\alpha$ is constant, so the optimal choice of the technology levels leads to a first-order condition that sets the capital share equal to the constant $\frac{\beta}{\alpha+\beta}$.

The constancy of the capital share then leads to two useful and interesting results. First, the global production function takes a Cobb-Douglas form: for any levels of the inputs K and L, and any location of the technology menu, N, the choice of technology leads the elasticity of output with respect to capital and labor to be constant.

In fact, it is easy to derive the exact form of the global production function by combining the local-global insights of Section 2.1 with the technology menu. For some technique i, recall the equivalent ways we have of describing the technique:

$$y_i \equiv a_i \tag{9}$$

$$k_i \equiv \frac{a_i}{b_i} \tag{10}$$

[4]In this case, the assumption that $\tilde{F}$ has an elasticity of substitution less than one means that the isoquant curves are more sharply curved than the technology menu, which has an elasticity of substitution equal to one. This guarantees an interior solution.

From the technology frontier in equation (8), we know that a_i and b_i are related by $a_i^\alpha b_i^\beta = N$. Simple algebra shows that y_i and k_i are therefore related by

$$y_i = (Nk_i^\beta)^{\frac{1}{\alpha+\beta}}. \tag{11}$$

That is, given the constant elasticity form of the technology frontier, a plot of the techniques in (k, y) space like that in Figure 1 yields a Cobb-Douglas production function. With this continuous formulation for the frontier, the global production function is equal to the technology frontier in (k, y) space.[5] Multiplying by L to get back to the standard form, the global production function is given by

$$Y = \left(NK^\beta L^\alpha\right)^{\frac{1}{\alpha+\beta}}. \tag{12}$$

That is, we get a Cobb-Douglas production function with constant returns to scale.

The second key result is related to the direction of technical change. To see this, consider embedding this production setup in a standard neoclassical growth model.[6] The fact that the global production function is Cobb-Douglas implies immediately that such a model will exhibit a balanced growth path with positive growth provided N grows exponentially.

The balanced growth path result turns out to have a strong implication for the direction of technical change. In particular, it implies that the level of b will be constant along the balanced growth path, and all growth will

[5]For this to be true, we need the local production techniques to paste up smoothly with the global production function. For example, if $\tilde{F}$ is a CES function with a capital share parameter λ (see, for example, equation (36) below), the global production function is actually *proportional* to that in equation (12). To make the factor of proportionality equal to one, we need the share parameter λ to equal $\frac{\beta}{\alpha+\beta}$, so that the factor share at $k = k_i$ is exactly $\frac{\beta}{\alpha+\beta}$.

[6]By this we mean the usual Ramsey-Cass-Koopmans model with isoelastic utility, constant population growth, and constant growth in N.

occur because a rises over time. To see this result, notice that the first order condition in equation (7) can be written as

$$\frac{bK\tilde{F}_1(bK, aL)}{aL\tilde{F}_2(bK, aL)} = \frac{\beta}{\alpha}. \tag{13}$$

Now, let $x \equiv bK/aL$. Because $\tilde{F}$ exhibits constant returns to scale, the marginal products are homogeneous of degree 0. This means we can rewrite equation (13) as

$$\frac{x\tilde{F}_1(x, 1)}{\tilde{F}_2(x, 1)} = \frac{\beta}{\alpha}. \tag{14}$$

Since x is the only variable in this equation, the optimal choice of technology is such that x is constant at all points in time.

Finally, we now need to show that along a balanced growth path, the only way $x \equiv bK/aL$ can be constant is if b is constant. Importantly, recall that output is always produced with some local production technique. That is,

$$Y_t = F(K_t, L_t; N_t) = \tilde{F}(b_t K_t, a_t L_t), \tag{15}$$

where b_t and a_t are the optimal choices of the technology levels. Because $\tilde{F}$ exhibits constant returns, we have

$$\frac{Y_t}{a_t L_t} = \tilde{F}\left(\frac{b_t K_t}{a_t L_t}, 1\right). \tag{16}$$

Since $x = bK/aL$ must be constant, this implies that Y/aL must also be constant. And this means that bK/Y must be constant. But we know that K/Y is constant along a balanced growth path in the neoclassical growth model, so this implies that b must be constant as well, which completes the proof. Moreover, the fact that the capital share equals $\frac{\beta}{\alpha+\beta}$ implies that the level of b is chosen so that the capital share is invariant to the capital-output ratio, one of the key results in Acemoglu (2003b).

Of course, the result that b must be constant along the balanced growth path is really just an application of the Steady-State Growth Theorem: If a neoclassical growth model exhibits steady-state growth with constant and positive factor shares, then either the production function is Cobb-Douglas or technical change is labor augmenting. In fact, we just proved a version of this theorem for the case in which the local production function is not Cobb-Douglas.[7]

What is the intuition for the result that technical change is purely labor augmenting? Since the local production function is not Cobb-Douglas, balanced growth requires bK/aL to be constant, so that bK and aL must grow at the same rate. In fact, since $Y = \tilde{F}(bK, aL)$, this suggests an alternative interpretation of the word "balanced" in the phrase "balanced growth path": the effective inputs bK and aL must be balanced in the sense that they grow at the same rate. But the only way this can happen is if b is constant. For example, we know that with b constant, K will grow at the same rate as aL. If b were to grow on top of this, bK would grow faster than aL, and growth would be unbalanced. The consequence that would result is that the factor shares would trend to zero and one.

In the context of our model, it is easy to be confused by this theorem. It is well-known that with Cobb-Douglas production, the "direction" of technical change has no meaning: capital-augmenting technical change can always be written as labor-augmenting. But the results just outlined seem to be that production is Cobb-Douglas and technical change is labor-augmenting. How can this be?

The key to resolving this confusion is to look back at equation (15). First, recall that production always occurs with some local production technique,

[7] For the proof of the general theorem, the classic reference is Uzawa (1961); see also Barro and Sala-i-Martin (1995) for a proof in the special case of factor-augmenting technologies. Jones and Scrimgeour (2004) present a formal statement of the theorem, discuss a version of Uzawa's proof, and develop intuition in the general case.

$\tilde{F}(b_t K_t, a_t L_t)$. Since this local technique has an elasticity of substitution less than one, the direction of technical change is a well-defined concept. Our result is that b_t is constant along a balanced growth path, so that technical change in the local production function is purely labor augmenting. Second, equation (15) also reminds us of the definition of the global production function, $F(K, L; N)$. It is this global production function that we show to be Cobb-Douglas. At any point in time, both "views" of the production function are possible, and it is by taking different points of view that we get our two results.

2.3. Discussion

We now pause to make some more general remarks about the baseline model. First, notice that an alternative way to set up the baseline model would be to write down the firm's full profit maximization problem. That is, in addition to choosing a and b, we could allow the firm to choose K and L, taking factor prices as given. We view the analysis of the global production function as conceptually coming a step before profit maximization. The global production function is defined over any combination of K and L; if one desires, one can embed this global production function into a model of how firms choose their inputs. For our purposes, however, all we are assuming about firm behavior is that they operate their technology efficiently. A helpful analogy might be that one can write down the cost-minimization problem as a precursor to the profit-maximization problem.[8]

[8] In the context of profit maximization, a more formal justification for the global production function approach can be given. For example, the full profit maximization problem can be written

$$\max_{a,K,L} \tilde{F}(H(a, N)K, aL) - wL - rK.$$

The global production function approach can be justified by noting that it is characterized by the first-order condition associated with the technology choice in the profit maximization problem.

Second, our problem is closely related to the world technology frontier problem considered by Caselli and Coleman (2004). Caselli and Coleman specialize to CES functions for both $\tilde{F}$ and the technology menu H and embed their setup in a profit maximization problem. They are concerned primarily with characterizing the choices of the technology levels in a cross-country context, rather than over time. But the similarity of the setups is interesting and suggests some potentially productive avenues for research.[9]

Finally, notice that the problem here is to choose the *levels* of a and b. Related problems appear in the literature on the direction of technical change; see Kennedy (1964), Samuelson (1965) and Drandakis and Phelps (1966). However, in these problems the choice variables and the constraints are typically expressed in terms of the growth rates of a and b rather than the levels. As Acemoglu (2003a) and others have pointed out, this results in an arbitrary optimization problem in the early literature related to maximizing the growth rate of output.

Acemoglu (2003b) recasts the traditional problem in terms of a 2-dimensional version of Romer (1990) with explicit microfoundations and profit maximizing firms. Under some strong — and arguably implausible[10] — conditions on the shape of the idea production functions, Acemoglu shows that technical change will be purely labor-augmenting in the long run and that the long-run capital share will be invariant to policies that change the

[9]Caselli and Coleman also contain a helpful discussion of the existence of interior versus corner solutions in their setup.

[10]The production functions for capital-ideas and labor-ideas must be parameterized "just so." In particular, let N denote the stock of labor-augmenting ideas. Then the cost of producing new labor-augmenting ideas relative to the cost of producing new capital-augmenting ideas must decline at exactly the rate $\dot{N}/N$. Plausible specifications — such as one in which the output good itself is the main input into the production of new ideas (in which case the relative cost of producing labor and capital ideas is constant) or the idea production function employed by Jones (1995) to remove scale effects from the growth rate (in which case the relative cost of producing labor ideas declines with N^{ϕ}) — lead to a model that does not exhibit a steady state with a positive capital share.

capital-output ratio. These results are obviously closely related to what we have here despite the considerably different approaches of the two papers. The main differences in terms of the results are that (a) we provide a very different perspective on the conditions needed to get technical change to be labor augmenting, and (b) we explicitly bring out the link to a Cobb-Douglas production function.[11]

To sum up, the insight from this baseline model is that if the technology frontier — i.e. the way in which the levels of a and b trade off — exhibits constant elasticities, then the global production function will be Cobb-Douglas and technological change will be labor-augmenting in the long run. But is there any reason to think that the technology frontier takes this particular shape?

3. MICROFOUNDATIONS: PARETO DISTRIBUTIONS

The baseline model is straightforward and yields strong predictions. However, it involves a very particular specification of the technology menu. It turns out that this specification can be derived from a model of ideas with substantially richer microfoundations. This is the subject of the current section.[12]

3.1. Setup

An idea in this economy is a technique for combining capital and labor to produce output. The production technique associated with idea i is $\tilde{F}(b_i K, a_i L)$. Because it results in a more tractable problem that yields

[11]The results here suggest that one might interpret Acemoglu's setup as providing a Cobb-Douglas production function in the long run. In contrast, our result delivers Cobb-Douglas production at any point in time.

[12]I owe a large debt to Sam Kortum in this section. A previous version of this paper contained a much more cumbersome derivation of the Cobb-Douglas result. Kortum, in discussing this earlier version at a conference, offered a number of useful comments that simplify the presentation, including the Poisson approach that appears in the Appendix.

analytic results, we make the extreme assumption that this local production technology is Leontief:

$$Y = \tilde{F}(b_i K, a_i L) = \min\{b_i K, a_i L\}. \tag{17}$$

Of course, the intuition regarding the global production function suggests that it is determined by the distribution of ideas, not by the shape of the local production function. In later simulation results, we confirm that the Leontief assumption can be relaxed.

A production technique is parameterized by its labor-augmenting and capital-augmenting parameters, a_i and b_i. To derive the Cobb-Douglas result, we make a strong assumption about the distribution of ideas:

ASSUMPTION 3.1. *The parameters describing an idea are drawn from independent Pareto distributions:*

$$\text{Prob}\,[a_i \leq a] = 1 - \left(\frac{a}{\gamma_a}\right)^{-\alpha}, \;\; a \geq \gamma_a > 0 \tag{18}$$

$$\text{Prob}\,[b_i \leq b] = 1 - \left(\frac{b}{\gamma_b}\right)^{-\beta}, \;\; b \geq \gamma_b > 0, \tag{19}$$

where $\alpha > 0$*,* $\beta > 0$*, and* $\alpha + \beta > 1$*.*[13]

With this assumption, the joint distribution of a_i and b_i satisfies

$$G(b,a) \equiv \text{Prob}\,[b_i > b, a_i > a] = \left(\frac{b}{\gamma_b}\right)^{-\beta} \left(\frac{a}{\gamma_a}\right)^{-\alpha}. \tag{20}$$

We specify this distribution in its complementary form because this simplifies some of the equations that follow.

[13]This last condition that the sum of the two parameters be greater than one is needed so that the mean of the Fréchet distribution below exists. On a related point, recall that for a Pareto distribution, the kth moment exists only if the shape parameter (e.g. α or β) is larger than k.

Let $Y_i(K, L) \equiv \tilde{F}(b_i K, a_i L)$ denote output using technique i. Then, since $\tilde{F}$ is Leontief, the distribution of Y_i is given by

$$\begin{aligned} H(\tilde{y}) \equiv \text{Prob}\,[Y_i > \tilde{y}] &= \text{Prob}\,[b_i K > \tilde{y}, a_i L > \tilde{y}] \\ &= G\left(\frac{\tilde{y}}{K}, \frac{\tilde{y}}{L}\right) \\ &= \gamma K^{\beta} L^{\alpha} \tilde{y}^{-(\alpha+\beta)}, \end{aligned} \tag{21}$$

where $\gamma \equiv \gamma_a^{\alpha} \gamma_b^{\beta}$. That is, the distribution of Y_i is itself Pareto.[14]

3.2. Deriving the Global Production Function

The global production function describes, as a function of inputs, the maximum amount of output that can be produced using any combination of existing production techniques. We have already made one simplification in our setup by limiting consideration to Leontief techniques. Now we make another by ignoring combinations of techniques and allowing only a single technique to be used at each point in time. Again, this is a simplifying assumption that allows for an analytic result, but it will be relaxed later in the numerical simulations.

Let N denote the total number of production techniques that are available, and assume that the N ideas are drawn independently. Then, we define the global production function:

DEFINITION 3.1. The global production function $F(K, L; N)$ is given as

$$F(K, L; N) \equiv \max_{i=1,\ldots,N} \tilde{F}(b_i K, a_i L) \tag{22}$$

Let $Y = F(K, L; N)$. Since the N draws are independent, the distribution of the global production function satisfies

$$\text{Prob}\,[Y \le \tilde{y}] = (1 - H(\tilde{y}))^N .$$

[14]Since $b_i \geq \gamma_b$ and $a_i \geq \gamma_a$, the support for this distribution is $\tilde{y} \geq \min\{\gamma_b K, \gamma_a L\}$.

$$= \left(1 - \gamma K^{\beta} L^{\alpha} \tilde{y}^{-(\alpha+\beta)}\right)^{N}. \tag{23}$$

Of course, as the number of ideas N gets large, this probability for any given level of $\tilde{y}$ goes to zero. So to get a stable distribution, we need to normalize our random variable somehow, in a manner analogous to that used in the Central Limit Theorem.

In this case, the right normalization turns out to involve z_N, where

$$z_N \equiv \left(\gamma N K^{\beta} L^{\alpha}\right)^{\frac{1}{\alpha+\beta}}. \tag{24}$$

In particular, consider

$$\begin{aligned} \text{Prob}\,[Y \leq z_N \tilde{y}] &= \left(1 - \gamma K^{\beta} L^{\alpha} (z_N \tilde{y})^{-(\alpha+\beta)}\right)^{N} \\ &= \left(1 - \frac{\tilde{y}^{-(\alpha+\beta)}}{N}\right)^{N}. \end{aligned} \tag{25}$$

Then using the standard result that $\lim_{N\to\infty}(1 - x/N)^N = exp(-x)$ for any fixed value of x, we have

$$\lim_{N\to\infty} \text{Prob}\,[Y \leq z_N \tilde{y}] = \exp(-\tilde{y}^{-(\alpha+\beta)}) \tag{26}$$

for $\tilde{y} > 0$. This distribution is known as a Fréchet distribution.[15]

Therefore

$$\frac{Y}{(\gamma N K^{\beta} L^{\alpha})^{1/\alpha+\beta}} \overset{a}{\sim} \text{Fréchet}(\alpha+\beta). \tag{27}$$

The global production function, appropriately normalized, converges asymptotically to a Fréchet distribution. This means that as N gets large, the production function behaves like

$$Y \approx \left(\gamma N K^{\beta} L^{\alpha}\right)^{\frac{1}{\alpha+\beta}} \epsilon \tag{28}$$

[15]This is a special case of the much more general theory of extreme values. For a more general theorem relevant to this case, see Theorem 2.1.1 of Galambos (1978), as well as Kortum (1997) and Castillo (1988).

where ϵ is a random variable drawn from a Fréchet distribution with shape parameter $\alpha + \beta$ and a scale parameter equal to unity.

Here, we have derived the Cobb-Douglas result as the number of ideas goes to infinity. We will show in the simulations that the approximation for a finite number of ideas works well. In addition, the appendix shows how to obtain the Cobb-Douglas result with a finite number of ideas under the stronger assumption that the arrival of ideas follows a Poisson process.

4. DISCUSSION

The result given in equation (28) is one of the main results in the paper. If ideas are drawn from Pareto distributions, then the global production function takes, at least as the number of ideas gets large, the Cobb-Douglas form. For any given production technique, a firm may find it difficult to substitute capital for labor and vice versa, leading the curvature of the production technique to set in quickly. However, when firms are allowed to switch between production technologies, the global production function depends on the distribution of ideas. If that distribution happens to be a Pareto distribution, then the production function is Cobb-Douglas.

We can now make a number of remarks. First, the exponent in the Cobb-Douglas function depends directly on the parameters of the Pareto search distributions. The easier it is to find ideas that augment a particular factor, the lower is the relevant Pareto parameter (e.g. α or β), and the lower is the exponent on that factor. Intuitively, better ideas on average reduce factor shares because the elasticity of substitution is less than one. Some additional remarks follow.

4.1. Relationship to the Baseline Model

The simple baseline model given at the beginning of this paper postulated a technology menu and showed that if this menu exhibited a constant elasticity, then one could derive a Cobb-Douglas global production function. The model with microfoundations based on Pareto distributions turns out to deliver a stochastic version of this technology menu.

In the model, the stochastic version of this menu can be seen by considering an iso-probability curve Prob $[b_i > b, a_i > a] \equiv G(b,a) = C$, where $C > 0$ is some constant. With the joint Pareto distribution, this iso-probability curve is given by

$$b^{\beta} a^{\alpha} = \frac{\gamma}{C}. \tag{29}$$

This iso-curve exhibits constant elasticities and shifts up as the probability C is lowered, analogous to an increase in N in the baseline model.

In terms of the baseline model, the Pareto distribution therefore delivers $\eta_a = \alpha$ and $\eta_b = \beta$, and we get the same form of the global production function: compare (12) and (28).

4.2. Houthakker (1955–1956)

The notion that Pareto distributions, appropriately "kicked," can deliver a Cobb-Douglas production function is a classic result by Houthakker (1955–1956). Houthakker considers a world of production units (e.g. firms) that produce with Leontief technologies where the Leontief coefficients are distributed across firms according to a Pareto distribution. Importantly, each firm has limited capacity, so that the only way to expand output is to use additional firms. Houthakker then shows that the aggregate production function across these units is Cobb-Douglas.

The result here obviously builds directly on Houthakker's insight that Pareto distributions can generate Cobb-Douglas production functions. The result differs from Houthakker's in several ways, however. First, Houthakker's result is an aggregation result. Here, in contrast, the result applies at the level of a single production unit (be it a firm, industry, or country). Second, the Leontief restriction in Houthakker's paper is important for the result; it allows the aggregation to be a function only of the Pareto distributions. Here, in contrast, the result is really about the shape of the global production function, looking across techniques. The local shape of the production function does not really matter. This was apparent in the simple baseline model given earlier, and it will be confirmed numerically in Section 6.

Finally, Houthakker's result relies on the presence of capacity constraints. If one wants to expand output, one has to add additional production units, essentially of lower "quality." Because of these capacity constraints, his aggregate production function is characterized by decreasing returns to scale. In the context of an idea model, such constraints are undesirable: one would like to allow the firm to take its best idea and use it for every unit of production. That is, one would like the setup to respect the nonrivalry of ideas and the replication argument for constant returns, as is true in the formulation here.[16]

4.3. Evidence for Pareto Distributions?

The next main comment is that Pareto distributions are crucial to the result. Is there any evidence that ideas follow a Pareto distribution?

Recall that the defining property of the Pareto distribution is that the conditional probability Prob $[X \geq \tau x \mid X \geq x]$ for $\tau > 1$ is independent of x. The canonical example of a Pareto distribution is the upper tail of the income distribution. Indeed, it was this observation that led Pareto to formulate the distribution that bears his name. Given that we observe an income larger than x, the probability that it is greater than $1.1x$ turns out

[16]Lagos (2004) embeds the Houthakker formulation in a Mortenson-Pissarides search model to provide a theory of total factor productivity differences. In his setup, firms (capital) match with labor and have a match quality that is drawn from a Pareto distribution. Capital is the quasi-fixed factor so that the setup generates constant returns to scale in capital and labor. Nevertheless, because each unit of capital gets its own Pareto draw, a firm cannot expand production by increasing its size at its best match quality.

to be invariant to the level of x, at least above a certain point. For example, Saez (2001) documents this invariance for the United States in 1992 and 1993 for incomes between \$100,000 and \$30 million.

Evidence of Pareto distributions has also been found for patent values, profitability, citations, firm size, and stock returns. First, it is worth noting that many of the tests in this literature are about whether or not the relevant variable obeys a Pareto distribution. That is, Pareto serves as a benchmark. In terms of findings, this literature either supports the Pareto distribution or finds that it is difficult to distinguish between the Pareto and the lognormal distributions. For example, Harhoff, Scherer and Vopel (1997) examine the distribution of the value of patents in Germany and the United States. For patents worth more than \$500,000 or more than 100,000 Deutsche Marks, a Pareto distribution accurately describes patent values, although for the entire range of patent values a lognormal seems to fit better. Bertran (2003) finds evidence of a Pareto distribution for ideas by using patent citation data to value patents. Grabowski (2002) produces a graph of the present discounted value of profits for new chemical entities by decile in the pharmaceutical industry for 1990–1994 that supports a highly-skewed distribution.

Lotka (1926), a classic reference on scientific productivity, shows that the distribution of scientific publications per author is Pareto. This result appears to have stood the test of time across a range of disciplines, even in economics, as shown by Cox and Chung (1991). It also applies to citations to scientific publications (Redner 1998). Huber (1998) looks for this result among inventors and finds some evidence that the distribution of patents per inventor is also Pareto, although the sample is small. Other evidence of Pareto distributions is found by Axtell (2001) for the size of firms in the United States, and by Gabaix, Gopikrishnan, Plerou and Stanley (2003) for the upper tail of stock returns. Finally, somewhat further afield, Pareto distributions are documented by Sornette and Zajdenweber (1999) for world movie revenues, and by Chevalier and Goolsbee (2004) for book sales. While by no means dispositive, this evidence of Pareto distributions for a wide range of economic variables that are certainly related to ideas is suggestive.

In addition to the direct evidence, there are also conceptual reasons to be open to the possibility that ideas are drawn from Pareto distributions. First, consider Kortum (1997). Kortum formulates a growth model where productivity levels (ideas) are draws from a distribution. He shows that this model generates steady-state growth only if the distribution has a Pareto upper tail. That is, what the model requires is that the probability of finding an idea that is 5 percent better than the current best idea is invariant to the level of productivity embodied in the current best idea. Of course, this is almost the very definition of a steady state: the probability of improving economy-wide productivity by 5 percent can't depend on the level of productivity. This requirement is satisfied only if the upper tail of the distribution is a power function, i.e. only if the upper tail is Pareto.

Additional insight into this issue emerges from Gabaix (1999). Whereas Kortum shows that Pareto distributions lead to steady-state growth, Gabaix essentially shows the reverse in his explanation of Zipf's Law for the size of cities. He assumes that city sizes grow at a common exponential rate plus an idiosyncratic shock. He then shows that this exponential growth generates a Pareto distribution for city sizes.[17]

[17] An important additional requirement in the Gabaix paper is that there be some positive lower bound to city sizes that functions as a reflecting barrier. Otherwise, for example, normally distributed random shocks results in a log-normal distribution of city cizes. Alternatively, if the length of time that has passed since each city was created is a random variable with an exponential distribution, then no lower bound is needed and one recovers the Pareto result. See Mitzenmacher (2003) for a direct discussion of these alternatives, as well as Cordoba (2003) and Rossi-Hansberg and Wright (2004).

The papers by Kortum and Gabaix suggest that Pareto distributions and exponential growth are really just two sides of the same coin. The result in the present paper draws out this connection further and highlights the additional implication for the shape of production functions. Not only are Pareto distributions necessary for exponential growth, but they also imply that the global production function takes a Cobb-Douglas form.

5. THE DIRECTION OF TECHNICAL CHANGE

The second main result of the paper is related to the direction of technical change. It turns out that this same setup, when embedded in a standard neoclassical growth model, delivers the result that technological change is purely labor augmenting in the long run. That is, even though the largest value of b_i associated with any idea goes to infinity, this Pareto-based growth model delivers the result that $a(t)$ grows on average while $b(t)$ is stationary.

To see this result, we first embed our existing setup in a standard neoclassical growth model. The production side of the model is exactly as specified in Section 3. Capital accumulates in the usual way, and we assume the investment rate s is a constant:

$$K_{t+1} = (1-\delta)K_t + sY_t, \quad \delta, s \in (0,1). \tag{30}$$

Finally, we assume the cumulative stock of ideas, N_t, grows exogenously at rate $g > 0$:

$$N_t = N_0 e^{gt}. \tag{31}$$

As in Jones (1995) and Kortum (1997), one natural interpretation of this assumption is that ideas are produced by researchers, so that g is proportional to population growth.[18]

[18]For example, one could have $\Delta N_{t+1} = R_t^{\lambda} N_t^{\phi}$, where R_t represents the number of researchers working in period t. In this case, if the number of researchers grows at a constant

For this model, we have already shown that the global production function is (either for N large or for finite N using the Poisson approach in the appendix):

$$Y_t = \left(\gamma N_t K_t^{\beta} L_t^{\alpha}\right)^{\frac{1}{\alpha+\beta}} \epsilon_t. \tag{32}$$

It is then straightforward to show that the average growth rate of output per worker y in the model in a stationary steady state is[19]

$$E[\log \frac{y_{t+1}}{y_t}] \approx g/\alpha. \tag{34}$$

The growth rate of output per worker is proportional to the rate of growth of research effort. The factor of proportionality depends only on the search parameter of the Pareto distribution for the labor-augmenting ideas. In particular, the easier it is to find higher a_i, the faster is the average rate of economic growth.

The fact that this growth rate depends on α but not on β is the first clue that there is something further to explore here: if it is easier to find better labor-augmenting ideas, the average growth rate is higher, but if it is easier to find better capital-augmenting ideas, the average growth rate is unaffected.

To understand this fact, it is helpful to look back at the local production function. Even though the global production function is Cobb-Douglas, production at some date t always occurs with some technique $i(t)$:

$$Y_t = \tilde{F}(b_{i(t)} K_t, a_{i(t)} L_t). \tag{35}$$

exponential rate, then the growth rate of N converges to a constant that is proportional to this population growth rate.

[19]Rewriting the production function in per worker terms, one has

$$\log \frac{y_{t+1}}{y_t} = \frac{1}{\alpha+\beta} \log \frac{N_{t+1}}{N_t} + \frac{\beta}{\alpha+\beta} \log \frac{k_{t+1}}{k_t} + \log \frac{\epsilon_{t+1}}{\epsilon_t}. \tag{33}$$

Taking expectations of this equation and equating the growth rates of y and k yields the desired result.

Now recall the Steady-State Growth Theorem discussed earlier: If a neoclassical growth model exhibits steady-state growth with a nonzero capital share, then either the production function is Cobb-Douglas or technical change is labor augmenting. In this case, the (local) production function is not Cobb-Douglas and we do have a (stationary) steady state. Exactly the same proof that we gave earlier for the baseline model in Section 2.2 applies. The implication is that technical change must be labor-augmenting in the long run. That is, despite the fact that $\max_i b_i \to \infty$ as $t \to \infty$, the time path for $b_{i(t)}$ — i.e. the time path of the b_i's associated with the ideas that are actually used — must have an average growth rate equal to zero in the limit. The intuition is also the same as in the simple baseline model: to keep the factor shares constant, growth must be balanced in the sense that bK and aL must grow at the same rate, and the only way this can happen is if b is stable.[20]

6. SIMULATION RESULTS

We now turn to a full simulation based on the Pareto model. In addition to providing an illustration of the results, we take this opportunity to relax the Leontief assumption on the local production function. Instead, we assume the local production function takes the CES form:

$$Y_t = \tilde{F}(b_i K_t, a_i L_t) = \left(\lambda(b_i K_t)^\rho + (1-\lambda)(a_i L_t)^\rho\right)^{1/\rho}, \qquad (36)$$

[20]This result leads to an important observation related to extending the model. Recall that with the Pareto assumption, γ_b is the smallest value of b that can be drawn, and similarly γ_a is the smallest value of a that can be drawn. Now consider allowing these distributions to shift. There seems to be no obstacle to allowing for exponential shifts in γ_a over time. However, increases in γ_b turn out to lower the capital share in the model. If γ_b were to rise exponentially, the capital share would be driven toward zero, on average. This does not, of course, mean that γ_b has never shifted historically; only that it should not have exhibited large shifts during the recent history when we have observed relatively stable factor shares. An alternative way in which the distributions may shift out over time is if the curvature parameters α and β shift. As long as the ratio α/β doesn't change, it may be possible to allow the mass of the distributions to shift out while keeping the capital share stable.

where $\rho < 0$ so that the elasticity of substitution is $\sigma \equiv \frac{1}{1-\rho} < 1$. We also allow production units to use two production techniques at a time in order to convexify the production set, analogous to the picture given at the beginning of the paper in Figure 1.

The remainder of the model is as specified before. Apart from the change to the CES function, the production setup is the same as that given in Section 3 and the rest of the model follows the constant saving setup of Section 5.

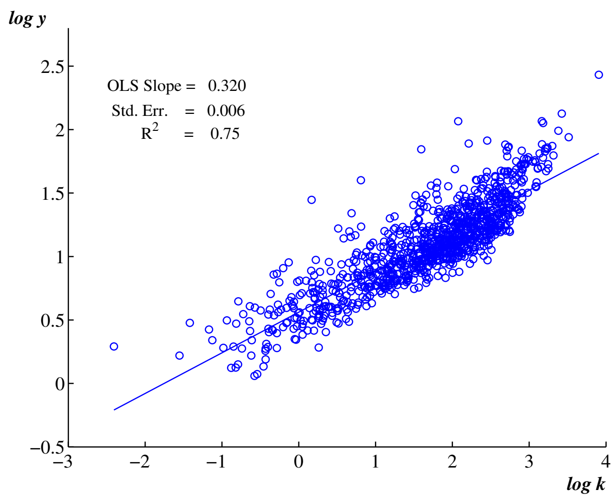
We begin by showing that the CES setup still delivers a Cobb-Douglas global production function, at least on average. For this result, we repeat the following set of steps to obtain 1000 capital-output pairs: We first set $N = 500$ so that there are 500 ideas in each iteration. We compute the convex hull of the CES functions associated with these ideas to get a global production function.[21] Next, we choose a level of capital per worker k randomly from a uniform distribution between the smallest value of k_i and the largest value of k_i for the iteration. Finally, we record the output of the global production function associated with this input.

Following this procedure yields a graph like that shown in Figure 3. The key parameter values in this simulation are $\alpha = 5$ and $\beta = 2.5$, so that the theory suggests we should expect a Cobb-Douglas production function with a capital exponent of $\beta/\alpha+\beta = 1/3$. As the figure shows, the relation between $\log y$ and $\log k$ is linear, with a slope that is very close to this value.

We next consider a simulation run for the full dynamic time path of the Pareto model. Continuing with the parameter choices already made, we additionally assume $g = .10$, which implies an annual growth rate of 2

[21] Computing the convex hull of the overlapping CES production functions is a computationally intensive problem, especially when the number of ideas gets large. To simplify, we first compute the convex hull of the (k_i, y_i) points. Then, we compute the convex hull of the CES functions associated with this limited set of points. To approximate the CES curve, we divide the capital interval into 100 equally-spaced points.

FIGURE 3. The Cobb-Douglas Result

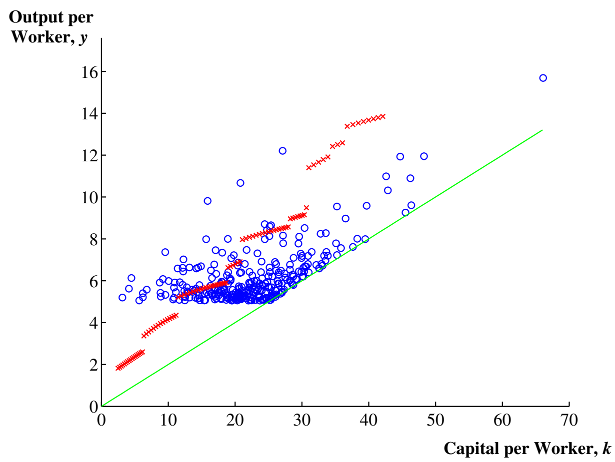


Note: The figure shows 1000 capital-output combinations from the global production function. The parameter values used in the simulation are $N = 500$, $\alpha = 5$, $\beta = 2.5$, $\gamma_a = 1$, $\gamma_b = 0.2$, and $\rho = -1$.

percent for output per worker in the steady state. We simulate this model for 100 years and plot the results in several figures.[22] Figure 4 shows a subset of the more than 1 million techniques that are discovered over these 100 periods. In particular, we plot only the 300 points with the highest values of y (these are shown with circles "o"). Without this truncation, the lower triangle in the figure that is currently blank but for the "x" markers is filled in as solid black. In addition, the capital-output combinations that are actually used in each period are plotted with an "x". When a single technique is used for a large number of periods, the points trace out the local CES production function. Alternatively, if the economy is convexifying by using two techniques, the points trace out a line. Finally, when the economy switches to a new technique, the capital-output combinations jump upward.

[22]Additional parameter values used in the simulation are listed in the notes to Figure 4.

FIGURE 4. Production in the Simulated Economy



Note: Circles indicate ideas, the "x" markers indicate capital-output combinations that are actually used. The model is simulated for 100 periods with $N_0 = 50$, $\alpha = 5$, $\beta = 2.5$, $g = .10$, $\gamma_a = 1$, $\gamma_b = 0.2$, $k_0 = 2.5$, $s = 0.2$, $\delta = .05$, and $\rho = -1$.

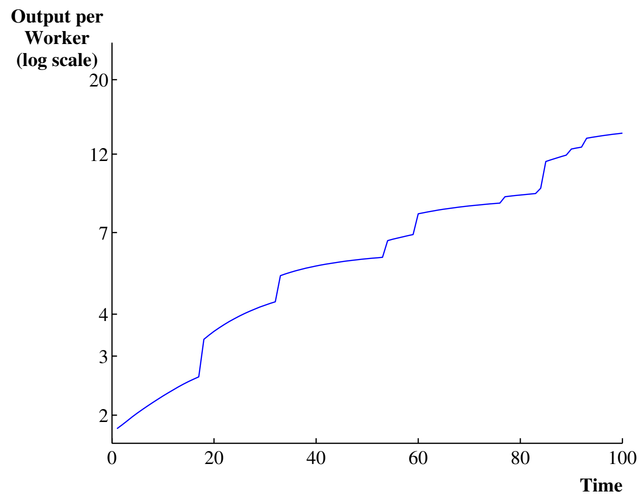
Figure 5 shows output per worker over time, plotted on a log scale. The average growth rate of output per worker in this particular simulation is 1.63 percent, as compared to the theoretical value of 2 percent implied by the parameter values, given by g/α.[23]

A feature of the model readily apparent in Figure 5 is that the economy switches from one production technique to another rather infrequently. These switches are shown in the graph as the jumps that occur roughly every 15 years or so. Moreover, when the jumps occur, they are typically quite large.

What explains these patterns? Recall that matching a Cobb-Douglas exponent on capital of 1/3 pins down the ratio of α/β, but it does not tell

[23]We compute the average growth rate by dropping the first 20 observations (to minimize the effect of initial conditions) and then regressing the log of output per worker on a constant and a time trend.

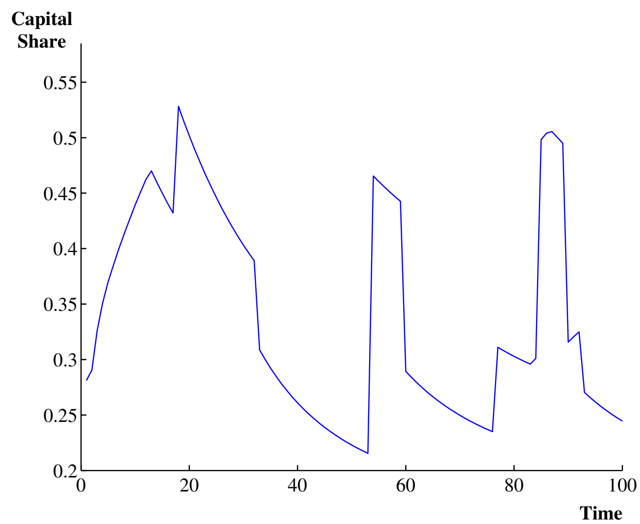
FIGURE 5. Output per Worker over Time



Note: See notes to Figure 4.

us the basic scale of these parameters. The studies cited earlier related to patent values, scientific productivity, and firm size typically find Pareto parameters that are in the range of 0.5 to 1.5. We have chosen higher values of $\alpha = 5$ and $\beta = 2.5$. The following exercise is helpful in thinking about this: What is the median value of a productivity draw, conditional on that draw being larger than some value, x? If α is the Pareto parameter, then the answer to this question turns out to be $2^{1/\alpha}x \approx (1 + 0.7/\alpha)x$. For example, if $\alpha = 1$, then the median value, conditional on a draw being higher than x, is $2x$. This says that the average idea that exceeds the frontier exceeds it by 100 percent! This implies very large jumps, which might be plausible at the micro level but seem too large at the macro level. A value of $\alpha = 5$ instead gives an average jump of about 14 percent, which is still somewhat large, and which explains the large jumps in Figure 5. We could have chosen an even larger Pareto parameter to yield smaller and more

FIGURE 6. The Capital Share over Time



Note: See notes to Figure 4.

frequent jumps, but this would have placed the value further from the range suggested by empirical studies. If the goal were to produce a simulation that could match the small, frequent jumps in the aggregate data with plausible Pareto coefficients, I suspect one would need a richer model that includes multiple sectors and firms. The jumps at the micro level would be large and infrequent, while aggregation would smooth things out at the macro level. This is an interesting direction for further research.[24]

Figure 6 plots the capital share $F_K K/Y$ over time. Even though the economy grows at a stable average rate, the capital share exhibits fairly large movements. When the economy is using a single production technique, the accumulation of capital leads the capital share to decline. Alternatively,

[24]Gabaix (2004) is related to this point. That paper shows that with a Pareto distribution of firm sizes and a Pareto parameter less than two, idiosyncratic shocks are smoothed out at a substantially slower rate than the standard central limit theorem suggests.

when the economy is using two techniques to convexify the production set, the marginal product of capital is constant, so the capital share rises smoothly.

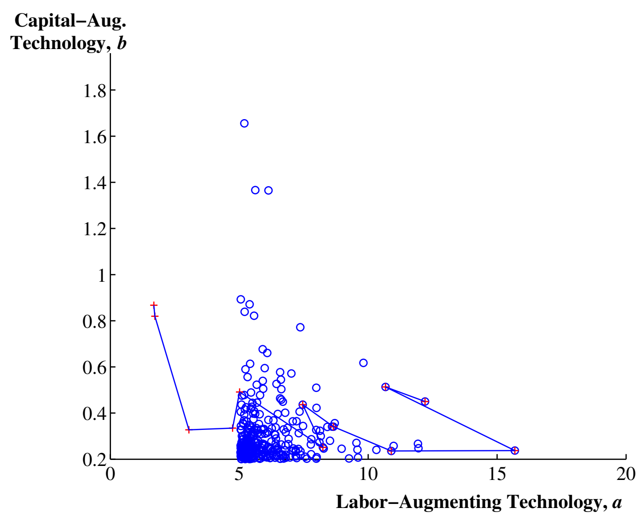
It is interesting to compare the behavior of the capital share in the Pareto model with the behavior that occurs in the simple baseline model. In the simple model, the economy equates the capital share to a function of the elasticity of the technology menu. If this elasticity is constant, then the capital share would be constant over time. Here, the technology menu exhibits a constant elasticity on average, but the menu is not a smooth, continuous function. Quite the opposite: the extreme value nature of this problem means the frontier is sparse, as the example back in Figure 1 suggests. This means the capital share will be stationary, but that it can move around, both as the economy accumulates capital and as it switches techniques.

Figure 7 shows the technology choices that occur in this simulation. As in Figure 4, the 300 ideas with the highest level of $y_i = a_i$ are plotted. This time, however, the (a_i, b_i) pair corresponding to each idea is plotted. The graph therefore shows the stochastic version of the technology menu. In addition, the figure plots with a "+" the idea combinations that are actually used as the economy grows over time. Corresponding to the theoretical finding earlier, one sees that the level of b_i^* appears stationary, while the level of a_i^* trends upward. On average, technological change is labor augmenting.

7. CONCLUSION

This paper provides microfoundations for the standard production function that serves as a building block for many economic models. An idea is a set of instructions that tells how to produce with a given collection of inputs. It can be used with a different mix of inputs, but it is not especially effective with the different mix; the elasticity of substitution in production

FIGURE 7. Technology Choices



Note: From more than 1 million ideas generated, the 300 with the highest level of a are plotted as circles. The figure also plots with a "+" the (a_i, b_i) combinations that are used at each date and links them with a line. When two ideas are used simultaneously, the idea with the higher level of output is plotted. See also notes to Figure 4.

is low for a given production technique. Instead, producing with a different input mix typically leads the production unit to switch to a new technique. This suggests that the shape of the global production function hinges on the distribution of available techniques.

Kortum (1997) examined a model in which productivity levels are draws from a distribution and showed that only distributions in which the upper tail is a power function are consistent with exponential growth. If one wants a model in which steady-state growth occurs, then one needs to build in a Pareto distribution for ideas. We show here that this assumption delivers two additional results. Pareto distributions lead the global production function to take a Cobb-Douglas form and produce a setup where technological change in the local production function is entirely labor augmenting in the long run.

There are several additional directions for research suggested by this approach. First, our standard ways of introducing skilled and unskilled labor into production involve production functions with an elasticity of substitution bigger than one, consistent with the observation that unskilled labor's share of income seems to be falling.[25] How can this view be reconciled with the reasoning here?

Second, the large declines in the prices of durable investment goods are often interpreted as investment-specific technological change. That is, they are thought of as increases in b rather than increases in a.[26] This is the case in Greenwood, Hercowitz and Krusell (1997) and Whelan (2003), and it is also implicitly the way the hedonic pricing of computers works in the National Income and Product Accounts: better computers are interpreted

[25] See Katz and Murphy (1992) and Krusell, Ohanian, Rios-Rull and Violante (2000), for example.

[26] This is loose. In fact, they are thought of as increases in a term that multiplies investment in the capital accumulation equation. Of course, for many purposes this is like an increase in b.

as more computers. The model in this paper suggests instead that b might be stationary, so there is a tension with this other work. Of course, it is not at all obvious that better computers are equivalent to more computers. Perhaps a better computer is like having two people working with a single computer (as in extreme programming). In this case, better computers might be thought of as increases in a instead. This remains an open question. Alternatively, it might be desirable to have microfoundations for a Cobb-Douglas production function that permits capital-augmenting technological change to occur in the steady state.

Finally, one might ask how the model relates to recent discussions about the behavior of capital shares. The literature is in something of a flux. For a long time, of course, the stylized fact has been that capital's share is relatively stable. This turns out to be true at the aggregate level for the United States and Great Britain, but it is not true at the disaggregated level in the U.S. or in the aggregate for many other countries. Rather, the more accurate version of the fact appears to be that capital's share can exhibit large medium term movements and even trends over periods longer than 20 years in some countries and industries.[27] This paper is somewhat agnostic about factor shares. As shown in Figure 6, the Pareto model predicts the capital share may vary over time, while of course the baseline model implied a constant capital share. However, there are many other determinants of capital shares left out of this model, including aggregation issues and wedges between marginal products and prices, so care should be taken in interpreting the model along this particular dimension.

[27]The recent papers by Blanchard (1997), Bentolila and Saint-Paul (2003), and Harrison (2003) discuss in detail the facts about capital and labor shares and how they vary. Gollin (2002) is also related; that paper argues that in the cross-section of countries, labor shares are more similar than rough data on employee compensation as a share of GDP suggests because of the very high levels of self-employment in many poor countries.

APPENDIX: AN ALTERNATIVE DERIVATION OF THE COBB-DOUGLAS RESULT

Here we show how to derive the Cobb-Douglas result for a finite number of ideas. The key to this stronger result is an assumption common in the growth literature: the assumption that the discovery of ideas follows a Poisson process.[26]

We now make the research process explicit. New ideas for production are discovered through research. A single research endeavor yields a number of ideas drawn from a Poisson distribution with a parameter normalized to one. In expectation, then, each research endeavor yields one idea. Let N denote the cumulative number of research endeavors that have been undertaken. Then the number of ideas, n, that have been discovered as a result of these N attempts is a random variable drawn from a Poisson distribution with parameter N. This additional layer is the only change to the model in Section 3.

For a given number of production techniques, the global production function $F(K, L; n)$ is

$$F(K, L; n) \equiv \max_{i \in \{0,\ldots,n-1\}} \tilde{F}(b_i K, a_i L). \tag{A.1}$$

As before, let Y_i denote production using technique i with a given amount of capital and labor. Then

$$\begin{aligned} \text{Prob}\,[Y_i > \tilde{y}] &= \text{Prob}\,[b_i K > \tilde{y}, a_i L > \tilde{y}] \\ &= G(\tilde{y}/K, \tilde{y}/L). \end{aligned} \tag{A.2}$$

The output level associated with the global production function is then distributed as

$$\text{Prob}\,[\max_i \{Y_i\} \le \tilde{y}] = (1 - G(\tilde{y}/K, \tilde{y}/L))^n \tag{A.3}$$

[26] For example, see Aghion and Howitt (1992).

At this point, we can use the nice properties of the Poisson distribution to make further progress. Recall that $n \sim Poisson(N)$, so as a function of the total number of research attempts, N, we have

$$\begin{aligned}
\text{Prob}\left[\max\{\ Y_i\ \} \le \tilde{y}\right] \\
&= \sum_{n=0}^{\infty} \frac{e^{-N} N^n}{n!} \left(1 - G(\tilde{y}/K, \tilde{y}/L)\right)^n \\
&= e^{-N} \sum_{n=0}^{\infty} \frac{\left(N(1 - G(\tilde{y}/K, \tilde{y}/L))\right)^n}{n!} \\
&= e^{-N} \cdot e^{N(1-G(\cdot))} \\
&= e^{-NG(\tilde{y}/K, \tilde{y}/L)}. \qquad \text{(A.4)}
\end{aligned}$$

For a general joint distribution function G, this last equation describes the distribution of the global production function when cumulative research effort is N.[27]

Now assume, as in the main text, the ideas are drawn from a joint Pareto distribution, so that

$$\text{Prob}\left[Y_i > \tilde{y}\right] = G(\tilde{y}/K, \tilde{y}/L) = \gamma K^{\beta} L^{\alpha} \tilde{y}^{-(\alpha+\beta)} \qquad \text{(A.5)}$$

Combining this result with equation (A.4), it is straightforward to show that the distribution of the output that can be produced with the global production function, given inputs of K and L, is

$$\text{Prob}\left[\max\{Y_i\} \le \tilde{y}\right] = e^{-\gamma N K^{\beta} L^{\alpha} \tilde{y}^{-(\alpha+\beta)}}, \qquad \text{(A.6)}$$

which is the Fréchet distribution.

Finally, taking expectations over this distribution, one sees that expected output, given N cumulative research draws and inputs K and L, is given

[27] See Proposition 2.1 in Kortum (1997) for this style of reasoning, i.e. for an approach that uses a Poisson process to get an exact extreme value distribution that is easy to work with rather than an asymptotic result. See also Johnson, Kotz and Balakrishnan (1994), pages 11 and 91–92.

by

$$E[Y] \equiv E[\max Y_i] = \mu \left(\gamma N K^{\beta} L^{\alpha}\right)^{\frac{1}{\alpha+\beta}} \tag{A.7}$$

where $\mu \equiv \Gamma(1-1/(\alpha+\beta))$ is a constant that depends on Euler's factorial function.[28]

One can also use the distribution in equation (A.6) to write the level of output as a random variable:

$$Y = \left(\gamma N K^{\beta} L^{\alpha}\right)^{\frac{1}{\alpha+\beta}} \epsilon \tag{A.8}$$

where ϵ is a random variable drawn from a Fréchet distribution with parameter $\alpha + \beta$. That is, we get the same result as in equation (28), but exactly for finite N rather than as an asymptotic approximation.

[28] Surprisingly few of the reference books on extreme value theory actually report the mean of the Fréchet distribution. For a distribution function $F(x) = exp(-((x - \lambda)/\delta)^{-\beta})$, Castillo (1988) reports that the mean is $\lambda + \delta\Gamma(1 - 1/\beta)$ for $\beta > 1$.

REFERENCES

Acemoglu, Daron, "Factor Prices and Technical Change: From Induced Innovation to Recent Debates," in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds., *Knowledge, Information, and Expectations in Modern Macroeconomics*, Princeton, N.J.: Princeton University Press, 2003.

—, "Labor- and Capital-Augmenting Technical Change," *Journal of European Economic Association*, March 2003, *1*, 1–37.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, *60* (2), 323–351.

Atkinson, Anthony B. and Joseph E. Stiglitz, "A New View of Technological Change," *Economic Journal*, September 1969, *79* (315), 573–578.

Axtell, Robert L., "Zipf Distribution of U.S. Firm Sizes," *Science*, September 2001, *293*, 1818–1820.

Barro, Robert J. and Xavier Sala-i-Martin, *Economic Growth*, McGraw-Hill, 1995.

Basu, Susanto and David N. Weil, "Appropriate Technology and Growth," *Quarterly Journal of Economics*, November 1998, *113* (4), 1025–1054.

Bentolila, Samuel and Gilles Saint-Paul, "Explaining Movements in the Labor Share," 2003. CEMFI mimeo.

Bertran, Fernando Leiva, "Pricing Patents through Citations," November 2003. University of Rochester mimeo.

Blanchard, Olivier J., "The Medium Run," *Brookings Papers on Economic Activity*, 1997, *1997* (2), 89–141.

Caballero, Ricardo J. and Mohamad L. Hammour, "Jobless Growth: Appropriability, Factor Substitution, and Unemployment," *Carnegie Rochester Conference Series on Public Policy*, 1998, *48*, 51–94.

Caselli, Francesco and Wilbur John Coleman, "The World Technology Frontier," April 2004. Harvard University mimeo.

Castillo, Enrique, *Extreme Value Theory in Engineering*, London: Academic Press, 1988.

Chevalier, Judith and Austan Goolsbee, "Price Competition Online: Amazon Versus Barnes And Noble," *Quantitative Marketing and Economics*, forthcoming 2004.

Cordoba, Juan Carlos, "On the Distribution of City Sizes," 2003. Rice University mimeo.

Cox, Raymond and Kee H. Chung, "Patterns of Research Output and Author Concentration in the Economics Literature," *Review of Economics and Statistics*, November 1991, *73* (4), 740–747.

Drandakis, E.M. and Edmund S. Phelps, "A Model of Induced Invention, Growth, and Distribution," *Economic Journal*, December 1966, *76* (304), 823–840.

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, August 1999, *114* (3), 739–767.

___, "Power Laws and the Granular Origins of Aggregate Fluctuations," July 2004. MIT mimeo.

___, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Eugene Stanley, "A Theory of Power Law Distributions in Financial Market Fluctuations," *Nature*, 2003, *423*, 267–270.

Galambos, Janos, *The Asymptotic Theory of Extreme Order Statistics*, New York: John Wiley & Sons, 1978.

Gilchrist, Simon and John C. Williams, "Putty Clay and Investment: A Business Cycle Analysis," *Journal of Political Economy*, October 2000, *108* (5), 928–960.

Gollin, Douglas, "Getting Income Shares Right," *Journal of Political Economy*, April 2002, *110* (2), 458–474.

Grabowski, Henry, "Patents and New Product Development in the Pharmaceutical and Biotechnology Industries," July 2002. Duke University mimeo.

Grandmont, Jean-Michel, "Distributions of Preferences and the 'Law of Demand'," *Econometrica*, January 1987, *55* (1), 155–161.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell, "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, June 1997, *87* (3), 342–362.

Harhoff, Dietmar, Frederic M. Scherer, and Katrin Vopel, "Exploring the Tail of Patented Invention Value Distributions," 1997. WZB Working Paper 97-27.

Harrison, Ann E., "Has Globalization Eroded Labor's Share? Some Cross-Country Evidence," May 2003. U.C. Berkeley mimeo.

Hildenbrand, Werner, "On the 'Law of Demand'," *Econometrica*, July 1983, *51* (4), 997–1020.

Houthakker, Henrik S., "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," *Review of Economic Studies*, 1955–1956, *23* (1), 27–31.

Huber, John C., “Cumulative Advantage and Success–Breeds–Sucess: the Value of Time Pattern Analysis,” *Journal of the American Society for Information Science*, 1998, *49* (5), 471–476.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 2*, Wiley Interscience, 1994.

Jones, Charles I., “R&D-Based Models of Economic Growth,” *Journal of Political Economy*, August 1995, *103* (4), 759–784.

___ and Dean Scrimgeour, “The Steady-State Growth Theorem: A Comment on Uzawa (1961),” October 2004. U.C. Berkeley mimeo.

Katz, Lawrence and Kevin Murphy, “Changes in Relative Wages, 1963–1987: Supply and Demand Factors,” *Quarterly Journal of Economics*, February 1992, *107* (1), 35–78.

Kennedy, Charles M., “Induced Bias in Innovation and the Theory of Distribution,” *Economic Journal*, September 1964, *74* (295), 541–547.

Kortum, Samuel S., “Research, Patenting, and Technological Change,” *Econometrica*, 1997, *65* (6), 1389–1419.

Krusell, Per, Lee Ohanian, Jose-Victor Rios-Rull, and Giovanni Violante, “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica*, September 2000, *68* (5), 1029–1053.

Lagos, Ricardo, “A Model of TFP,” January 2004. NYU working paper.

Lotka, A.J., “The Frequency Distribution of Scientific Productivity,” *Journal of the Washington Academy of Sciences*, 1926, *16*, 317–323.

Mitzenmacher, Michael, “A Brief History of Generative Models for Power Law and Lognormal Distributions,” *Internet Mathematics*, forthcoming 2003, *1* (2).

Redner, Sidney, “How Popular is Your Paper? An Empirical Study of the Citation Distribution,” *European Physical Journal B*, 1998, *4*, 131–134.

Robinson, Joan, “The Production Function and the Theory of Capital,” *Review of Economic Studies*, 1953–1954, *21* (2), 81–106.

Romer, Paul M., “Endogenous Technological Change,” *Journal of Political Economy*, October 1990, *98* (5), S71–S102.

Rossi-Hansberg, Esteban and Mark L. J. Wright, “Urban Structure and Growth,” 2004. Stanford University mimeo.

Saez, Emmanuel, “Using Elasticities to Derive Optimal Tax Rates,” *Review of Economic Studies*, 2001, *68*, 205–229.

Samuelson, Paul A., “A Theory of Induced Innovations Along Kennedy-Weisacker Lines,” *Review of Economics and Statistics*, 1965, *47*, 343–356.

Sornette, Didier and Daniel Zajdenweber, "The economic return of research: the Pareto law and its implications," *European Physical Journal B*, 1999, *8* (4), 653–664. http://xxx.lanl.gov/abs/cond-mat/9809366.

Uzawa, Hirofumi, "Neutral Inventions and the Stability of Growth Equilibrium," *Review of Economic Studies*, February 1961, *28* (2), 117–124.

Whelan, Karl, "A Two-Sector Approach to Modeling U.S. NIPA Data," *Journal of Money, Credit, and Banking*, August 2003, *35* (4), 627–656.