Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail

Charles I. Jones

Stanford University and National Bureau of Economic Research

As Romer and Weitzman emphasized in the 1990s, new ideas are often combinations of existing ideas, an insight absent from recent models. In Kortum's research around the same time, ideas are draws from a probability distribution, and Pareto distributions play a crucial role. Why are combinations missing, and do we really need such strong distributional assumptions to get exponential growth? This paper demonstrates that combinatorially growing draws from standard thin-tailed distributions lead to exponential growth; Pareto is not required. More generally, it presents a theorem linking the max extreme value to the number of draws and the shape of the upper tail for probability distributions.

I. Introduction

It has long been appreciated that new ideas are often combinations of existing goods or ideas. Gutenberg's printing press was a combination of movable type, paper, ink, metallurgical advances, and a wine press. State-of-the-art photolithographic machines for making semiconductors weigh 180 tons and combine inputs from 5,000 suppliers, including robotic arms and mirrors of unimaginable smoothness (*Economist* 2020). Romer (1993)

I am grateful to Daron Acemoglu, Andy Atkeson, Pablo Azar, Ian Dew-Becker, Sebastian Di Tella, Brian Cevallos Fujiy, Guido Imbens, Pete Klenow, Sam Kortum, Lihua Lei, Erzo Luttmer, Ben Moll, Jeremy Pearce, Chris Tonetti, Akos Valentinyi, Tom Winberry, and numerous seminar participants for helpful discussions and comments and to Jack Hirsh for excellent research assistance. This paper was edited by Andrew Atkeson.

Electronically published June 23, 2023

Journal of Political Economy, volume 131, number 8, August 2023.

https://doi.org/10.1086/723631

observes that ingredients from a children's chemistry set can create more distinct combinations than there are atoms in the universe. Building on this insight, Weitzman (1998) constructs a growth model in which new ideas are combinations of old ideas. Because combinatorial growth is so fast, however, he finds that growth is constrained by our limitations in processing an exploding number of ideas, and the combinatorics plays essentially no formal role in determining the growth rate: there are so many potential combinations that the number is not a constraint. It is somewhat disappointing and puzzling that the combinatorial process does not play a more central role.

A separate literature highlights the links between exponential growth and Pareto distributions. In particular, Kortum (1997) introduced a new way of modeling economic growth and argued that Pareto distributions are crucial: if productivity is the maximum over a number of draws from a distribution (you use only the best idea), then exponential growth in productivity in his setup requires that the number of draws grows exponentially and that the distribution being drawn from is Pareto, at least in the upper tail. It is interesting that such a strong distributional assumption seems to be required. Perhaps the underlying distribution from which ideas are drawn is Pareto, but why would that be the case? After all, in many other applications in economics, the Pareto distribution is derived rather than assumed. For example, Gabaix (1999), Luttmer (2007), and Jones and Kim (2018) highlight that city sizes, firm employment, incomes, and wealth all feature Pareto distributions. However, that literature shows how these Pareto distributions emerge as an endogenous outcome. This raises the question of whether the Pareto distribution really is necessary in the Kortum approach. And regardless, what happened to the Romer and Weitzman insight that combinatorics should be central to understanding growth?

This paper answers these questions by combining the insights of Kortum (1997) and Weitzman (1998). Suppose that ideas are combinations of existing ingredients, much like a recipe. Each recipe has a productivity that is a draw from a probability distribution. As in Romer and Weitzman, the number of combinations we can create from existing ingredients is so astronomically large as to be essentially infinite, and we are limited by our ability to process these combinations. Let N_t denote the number of ingredients whose recipes have been evaluated as of date t. In other words, our "cookbook" includes all the possible recipes that can be formed from N_t ingredients: if each ingredient can be either included or excluded from a recipe, a total of 2^{N_t} recipes are in the cookbook. Finally, research consists of adding new recipes to the cookbook—that is, evaluating them and learning their productivities. In particular, suppose that researchers add new ingredients to the cookbook and learn their productivities in such a way that N_t grows exponentially. We call a setup with 2^{N_t} recipes with exponential growth in N_t *combinatorial growth.* (In the model below, the exponential growth in N_t will occur because of population growth in the number of researchers.)

One key result in the paper is this: combinatorial growth is so fast that drawing from a conventional thin-tailed distribution—such as the normal, exponential, or Weibull distribution—generates exponential growth in the productivity of the best recipe in the cookbook. Combinatorics and thin tails lead to exponential growth.

The way we derive this result leads to broader insights. For example, let K denote the cumulative number of draws (e.g., the number of recipes in the cookbook), and let Z_K be max of the K productivities. Let $\bar{F}(x)$ denote the probability that a draw has a productivity higher than x—the complement of the cumulative distribution function (cdf)—so that it characterizes the search distribution. Then a key condition derived below relates the rise in Z_K to the number of draws and to the search distribution: Z_K increases asymptotically so as to stabilize $K\bar{F}(Z_K)$. That is, given a time path for the number of draws K_t, the maximum productivity marches down the upper tail of the distribution so as to make $K_t\bar{F}(Z_{Kt})$ stationary.

Kortum (1997) can be viewed in this context: exponential growth in the max Z_K is achieved by an exponentially growing number of draws K from a Pareto tail in $\bar{F}(\cdot)$. Alternatively, with thinner-tailed distributions like the normal or the exponential, combinatorial growth in K is required to get exponential growth in the max. Even the Romer (1990) model can be viewed in this light: if we get a constant number of draws each period, the linear growth in K requires a log-Pareto tail for the search distribution if the max is to exhibit exponential growth.

Finally, the model features an important and testable empirical prediction. Kortum (1997) predicts that the flow of valuable new ideas should be constant over time, even as the number of researchers grows. For example, the discovery of 40,000 valuable new ideas in 1915, 1950, and 1985 can deliver constant exponential growth. The reason is that successful new ideas are "large" in some sense. They are drawn from a Pareto distribution and therefore generate proportional improvements in productivity on average. In the combinatorial version in which ideas are drawn from a thin-tailed distribution, new ideas are "small," and exponential growth therefore requires an exponentially rising flow of valuable new ideas. Empirical evidence suggests that the annual flows of patents, breakthrough patents, and academic publications have risen substantially over time, supporting the combinatorial model.

The remainder of the paper is organized as follows. After a brief literature review, section II explains the basic insights in a simple setting, while section III embeds the setup in a full growth model. Section IV connects our results with the literature on extreme value theory and shows how the results generalize to different distributions. Section V presents the evidence on patents and publications, providing empirical support for the model. Section VI then suggests a number of important directions for future research, including the distributions of markups and productivity, the frequency of creative destruction, and broader applications of the combinatorial approach to technology diffusion, international trade, and search models. This paper develops a new tool, and it would be valuable to apply that tool in many existing literatures.

Literature review.—Beyond Kortum (1997) and Weitzman (1998), the most important inspiration for this paper is Acemoglu and Azar (2020). They study endogenous production networks in which every good uses a combination of other goods as an intermediate input. If there are N goods in the economy, then there are 2^N possible combinations of intermediate goods that could be used to produce a particular product, and Acemoglu and Azar (2020) let the productivity of each of these combinations be a draw from a probability distribution. Their setup inspired the approach taken in this paper.

The two papers differ in thinking about how the number of goods/ingredients evolves over time. Because it is not the main contribution of their paper, Acemoglu and Azar (2020) focus on the case in which one new good gets introduced each period, so there is arithmetic growth in N_t and therefore exponential growth in 2^{N_t}. For this to produce exponential growth in productivity, they require the standard Kortum (1997) assumption that the probability distribution determining productivity has a Pareto upper tail.[1] Their corollary 2 suggests that broader results are possible, and the present paper can be viewed as exploring those broader results.

Akcigit, Kerr, and Nicholas (2013) provide the best empirical evidence to date of the importance of combinations to idea production. The share of US patents based on a novel combination of existing technologies rose from 50% in the 1800s to more than 75% in recent decades. Interestingly, they also show that more than 80% of technologies ever developed still get incorporated into new patents today, suggesting an incredibly long impact of past technologies on today's innovation. Fleming and Sorenson (2001) and Youn et al. (2015) also provide evidence from patent data on the importance of combinations.

Agrawal, McHale, and Oettl (2019) build a growth model based explicitly on combinations in the idea production function. They assume that the elasticity of new ideas with respect to combinations declines to zero in order to prevent explosive growth.

[1] They state the assumption in a different form: that the log of productivity is drawn from a Gumbel distribution. But, as they note, this is identical to saying that productivity itself is drawn from a Fréchet distribution.

II. Combining Weitzman and Kortum

Suppose that there are a huge number of ingredients that can potentially be combined into recipes, which we call ideas. Moreover, new ideas can also serve as future ingredients, making the number of potential combinations effectively infinite. Our cookbook, $\mathcal{C}$ is the set of all recipes we have evaluated as of some point in time. Let K denote the number of recipes in the cookbook.

Each recipe can be good or bad or somewhere in between. In one of the early seminars in which Paul Romer discussed these combinatorial calculations, George Akerlof is said to have remarked, "Yes, the number of possible combinations is huge, but aren't most of them like chicken ice cream!" Suppose that the productivity associated with each recipe is an independent draw from some distribution. In particular, let z_c denote the productivity of recipe c, and let $F(x)$ be the cdf for each independent z_c.

Now assume that we are interested in only the best recipe in our cookbook. That is, different ideas have different productivities, z_c, and we use the idea with the highest productivity, as in Kortum (1997). Let $Z_K \equiv \max z_c$, where $c \in \{1, \dots, K\}$. Because we care about the best idea, it is convenient to define the tail cdf:

$$\bar{F}(x) \equiv \Pr[z_c \geq x] = 1 - F(x). \tag{1}$$

From a growth perspective, the question is this: How does the productivity associated with the best idea, Z_K, change as the number of recipes in the cookbook, K, increases over time? And in particular, under what conditions can we get exponential growth in Z_K?

To answer these questions, consider the distribution of the maximum productivity, Z_K, if we have taken K draws from the distribution $F(x)$. Because the draws are independent,

$$\begin{aligned}\Pr[Z_K \leq x] &= \Pr[z_1 \leq x, z_2 \leq x, \dots, z_K \leq x] \\ &= F(x)^K \\ &= (1 - \bar{F}(x))^K.\end{aligned} \tag{2}$$

If we take more and more draws from the distribution over time so that K gets larger, then obviously $F(x)^K$ shrinks. To get a stable distribution, we need to normalize the max by some function of K, analogous to how in the central limit theorem we multiply the mean by the square root of the number of observations to get a stable distribution. Mechanically, if we replace the $\bar{F}(x)$ on the right-hand side of (2) with something that depends on $1/K$ and then take the limit as K goes to infinity, the exponential function will appear.

The following theorem provides a general result that will be useful in our growth application but may be useful more broadly as well.

THEOREM 1 (A simple extreme value result). Let Z_K denote the maximum value from $K > 0$ independent draws from a continuous distribution $F(x)$, with $\bar{F}(x) \equiv 1 - F(x)$ strictly decreasing on its support. Then for $m \geq 0$,

$$\boxed{\lim_{K\to\infty} \Pr[K\bar{F}(Z_K) \geq m] = e^{-m}.} \tag{3}$$

Proof. Given that Z_K is the max over K independently and identically distributed (i.i.d.) draws, we have

$$\Pr[Z_K \leq x] = (1 - \bar{F}(x))^K. \tag{4}$$

Let $M_K \equiv K\bar{F}(Z_K)$ denote a new random variable. Then for $0 \leq m < K$,

$$\begin{aligned}\Pr[M_K \geq m] &= \Pr[K\bar{F}(Z_K) \geq m]\\ &= \Pr\left[\bar{F}(Z_K) \geq \frac{m}{K}\right]\\ &= \Pr\left[Z_K \leq \bar{F}^{-1}\left(\frac{m}{K}\right)\right]\\ &= \left(1 - \frac{m}{K}\right)^K,\end{aligned}$$

where the penultimate step uses the fact that $\bar{F}(x)$ is a strictly decreasing and continuous function and the last step uses the result from (4). The fact that $\lim_{K\to\infty}(1 - m/K)^K = e^{-m}$ proves the result. QED

Let us pause here to notice what is happening in theorem 1. We have a new random variable, $K\bar{F}(Z_K)$. As K goes to infinity, Z_K—the max over K draws from the distribution—is getting larger. So $\bar{F}(Z_K)$—the probability the next draw exceeds Z_K—is shrinking toward zero as we march down the tail of the distribution. Multiplying by K raises the value away from zero, and it is the product $K\bar{F}(Z_K)$ that is asymptotically stationary. Theorem 1 says that under very weak conditions—basically, that the underlying distribution we draw from is continuous and monotone—$K\bar{F}(Z_K)$ converges in distribution to a standard exponential distribution.

A few remarks about this theorem are helpful. First, for using the theorem, it is convenient to note that the result can be written as

$$K\bar{F}(Z_K) = \varepsilon + o_p(1), \tag{5}$$

where ε is an exponential random variable with a mean equal to 1. This expression can be rearranged to provide intuition for the result:

$$\bar{F}(Z_K) \equiv \Pr[\text{Next draw} > Z_K] = \frac{1}{K}(\varepsilon + o_p(1)). \tag{6}$$

That is, as K gets large and apart from some random variation, the probability that the next draw exceeds the max is $1/K$, regardless of the distribution we are drawing from. This ties in nicely with the theory of record breaking (Glick 1978). For example, suppose that we have observed K days of temperatures in Palo Alto. What is the probability that today's temperature is a record high? If everything is i.i.d., then the unconditional answer is just $1/K$: any day could be the record. Theorem 1 shows that this intuition carries over even for the conditional probability; the randomness associated with ε captures the fact that we may have a particularly high or low maximum after K draws relative to what one would usually expect.

Next, nothing in the theorem requires that the distribution be unbounded. For example, the theorem applies to the uniform distribution as well: even though the max is bounded, $\bar{F}(Z_K)$ is falling to zero, and blowing this up by the factor K leads to an exponential distribution for the product.

Finally, an alternative version of theorem 1 is presented in section III that uses a Poisson assumption as in Kortum (1997) to derive a similar result at each point in time without needing to take the limit as t goes to infinity.

Results related to theorem 1 are of course known in the mathematical statistics literature. The earliest reference I have found is Barton and David (1959). It is also closely related to proposition 3.1.1 in Embrechts, Mikosch, and Klüppelberg (1997). Galambos (1978, chap. 4) develops a "weak law of large numbers" and a "strong law of large numbers" for extreme values; some of the results below will fit this characterization.[2] However, the tight link between the number of draws, the shape of the tail, and the way the maximum increases is not emphasized in these treatments. More generally, I discuss the result's relationship with standard extreme value theory in section IV.

The result in (3) means that $K\bar{F}(Z_K)$ is asymptotically stationary. Since Z_K and K are both rising, the rate at which the tail $\bar{F}(\cdot)$ decays tells us how the rates of increase of Z_K and K are related. We now apply this logic to growth models, first as in Kortum (1997) and then in a new way involving combinatorics.

A. *Kortum (1997)*

Kortum (1997) showed one way to get exponential growth in productivity Z_K in a setup similar to this: assume that $F(x)$ is a Pareto distribution, at least in the upper tail, and have K grow exponentially—for example, because of population growth in the number of researchers.

[2] But not all: e.g., the Kortum (1997) result and the Romer (1990) example at the end are convergence in distribution results, not convergence in probability results.

To see how this works, let $F(x) = 1 - x^{-\beta}$ so that $\bar{F}(x) = x^{-\beta}$, which is a Pareto distribution where a higher β means a thinner upper tail. In this case, $K\bar{F}(Z_K) = KZ_K^{-\beta}$, and theorem 1 gives

$$\begin{aligned} K\bar{F}(Z_K) &= \varepsilon + o_p(1), \\ KZ_K^{-\beta} &= \varepsilon + o_p(1), \\ \frac{K}{Z_K^{\beta}} &= \varepsilon + o_p(1), \end{aligned}$$

and therefore

$$\boxed{\frac{Z_K}{K^{1/\beta}} = (\varepsilon + o_p(1))^{-1/\beta}.} \tag{7}$$

In words, to get a stable distribution for the max over K draws from a Pareto distribution, we divide the max Z_K by $K^{1/\beta}$. This scaled-down max then is distributed asymptotically just like $\tilde{\varepsilon} \equiv \varepsilon^{-1/\beta}$, which has a Fréchet distribution. If the number of draws K grows exponentially at rate g_K (say, because each researcher gets one draw per period and there is population growth), then the growth rate of productivity Z_K asymptotically averages to

$$g_Z = \frac{g_K}{\beta}. \tag{8}$$

It equals the growth rate of the number of draws deflated by β, the rate at which good ideas are getting harder to find. This is the Kortum (1997) result.

B. *Weitzman Meets Kortum*

The Kortum result is beautiful, and it may be the way the world works. However, there are two features that are slightly uncomfortable. First, does the real world's idea distribution have a Pareto upper tail? Maybe. But given the large literature on generating Pareto distributions from exponential growth, it is slightly uncomfortable to have to assume an underlying Pareto distribution to get economy-wide growth. Can we do without this assumption?

Second, the combinatorics of ideas that Romer (1993) and Weitzman (1998) emphasized is entirely missing from this structure. What we show next is that addressing these two concerns together reveals an elegant alternative.

Let us change the Kortum setup in two ways. First, rather than drawing from a distribution with a Pareto upper tail, we draw from a standard thin-tailed distribution, such as the normal or exponential. To illustrate the logic, we begin with the exponential distribution: $F(x) = 1 - e^{-\theta x}$ so that $\bar{F}(x) = e^{-\theta x}$.

Second, let us assume that our cookbook consists of all recipes that come from combining N ingredients. Each ingredient can be either included or excluded from a recipe, so there are a total of $K = 2^N$ recipes. (Recall that $2^N = \Sigma_{k=0}^{N} \binom{N}{k}$, the total number of combinations.) The economy picks from $K = 2^N$ different combinations and chooses the recipe with the highest productivity. We say that K exhibits combinatorial growth if $K = 2^N$ and N itself grows at a constant and positive exponential rate.

Applying theorem 1 to this setup with $\bar{F}(x) = e^{-\theta x}$ gives

$$\begin{aligned} K\bar{F}(Z_K) &= \varepsilon + o_p(1), \\ Ke^{-\theta Z_K} &= \varepsilon + o_p(1), \\ \Rightarrow \log K - \theta Z_K &= \log(\varepsilon + o_p(1)), \\ \Rightarrow Z_K &= \frac{1}{\theta}[\log K - \log(\varepsilon + o_p(1))], \\ \Rightarrow \frac{Z_K}{\log K} &= \frac{1}{\theta}\left(1 - \frac{\log(\varepsilon + o_p(1))}{\log K}\right), \end{aligned} \tag{9}$$

and therefore

$$\boxed{\frac{Z_K}{\log K} \xrightarrow{p} \text{Constant},} \tag{10}$$

where here and later we will follow the convention that "Constant" denotes an unimportant positive constant that may change across equations. With draws from an exponential distribution, the max grows asymptotically with the natural log of the number of draws, a well-known result.

If the number of draws K were to grow exponentially at rate g_K, then productivity would grow linearly rather than exponentially, and the exponential growth rate would converge to zero, a point noted by Kortum (1997).

A key insight here is that if the number of draws is combinatorial instead, exponential growth is restored. In particular, if $K = 2^N$ and N grows exponentially at rate g_N, then

$$\frac{Z_K}{\log K} = \frac{Z_K}{N \log 2} \xrightarrow{p} \text{Constant}, \tag{11}$$

and the asymptotic growth rate of productivity in this economy will equal

$$g_Z = g_{\log K} = g_N. \tag{12}$$

Productivity growth is asymptotically equal to the growth rate of the number of ingredients whose recipes have been evaluated.

To summarize, the first new growth result is this: if recipes are combinations of N ingredients, and if the number of ingredients processed by the economy grows exponentially over time, then we no longer require draws from a thick-tailed Pareto distribution. Combinatorial expansion is so fast that we get enough draws from the thin-tailed exponential distribution to generate exponential growth in productivity.

C. *The Weibull Distribution*

A convenient shortcut allows us to generalize this result to other distributions. For now, we show how it generalizes to the Weibull distribution, as this will be particularly useful. In section IV, we will derive a necessary and sufficient condition for combinatorial draws to generate exponential growth, precisely characterizing the generality.

Equation (10) states that the max from K draws of an exponential, divided by $\log K$, converges in probability to a constant. Now, consider the Weibull distribution, $F(x) = 1 - e^{-x^\beta}$, and define $y = x^\beta$. If x is distributed as Weibull, then y is exponentially distributed. We can combine this change of variables with the scaling result for an exponential:

$$\begin{aligned} \frac{\max y}{\log K} &\xrightarrow{p} \text{Constant} \\ \Rightarrow \frac{\max x^\beta}{\log K} &\xrightarrow{p} \text{Constant} \\ \Rightarrow \frac{\max x}{(\log K)^{1/\beta}} &\xrightarrow{p} \text{Constant}. \end{aligned} \tag{13}$$

That is, the maximum over K draws from a Weibull distribution grows asymptotically as $(\log K)^{1/\beta}$. Assuming $K = 2^N$, the max grows with $N^{1/\beta}$, and if N grows exponentially at rate g_N, the growth rate of the max is asymptotically given by

$$g_Z^{\text{weibull}} = \frac{g_N}{\beta}. \tag{14}$$

Intuitively, a higher value of β means a thinner tail of the Weibull distribution—the exponential tail decays more rapidly. The growth rate of the max is the growth rate of the number of ingredients deflated by β, the rate at which ideas are getting harder to find. The Weibull distribution is to combinatorial growth what the Pareto distribution was to an exponentially growing number of draws in Kortum (1997).

III. Growth Model

This section embeds the extreme value logic provided above into a basic growth model. The setup is similar to Kortum (1997) except that we use a thin-tailed search distribution and combinatorial growth in the number of draws.

A. A Poisson Version of Theorem 1

We first state a corollary to theorem 1 that uses a Poisson assumption to get the extreme value result for all t rather than as an asymptotic result. I am grateful to Sam Kortum for suggesting it and outlining a derivation.

COROLLARY 1 (Poisson version of theorem 1). Let Z_K denote the maximum over P independent draws from a distribution with tail cdf $\bar{F}(x)$ that is strictly decreasing and continuous on its support, and suppose that P is distributed as Poisson with parameter K. Then for $0 \leq m < K$ and when $P > 0$ (so there are observations over which to take the max),

$$\Pr[K\bar{F}(Z_K) \geq m] = \frac{e^{-m} - e^{-K}}{1 - e^{-K}}. \tag{15}$$

Proof.—See the appendix.

In the corollary, notice that the e^{-K} term appears because $\Pr[P = 0] = e^{-K}$ and $\Pr[P > 0] = 1 - e^{-K}$—the max exists only once $P > 0$. Also, notice that as $K \to \infty$, we get the pure exponential distribution, as in theorem 1. The advantage of this Poisson version is that it applies for any K, not just asymptotically. Therefore, we can average over a continuum of sectors to get rid of the randomness and then use continuous-time methods for the growth theory, which simplifies the presentation.

B. The Environment

The economic environment for the full growth model is shown in table 1. The setup embeds combinatorial draws from a Weibull distribution into a continuous-time growth framework.

Aggregate output is a constant elasticity of substitution (CES) combination of a unit measure of varieties, as in equation (16). The production of each variety is given by (17). Each variety is produced using a different recipe from the cookbook. A recipe uses $M_{it} \leq N_t$ ingredients that combine in a CES fashion, and one unit of each ingredient can be produced with one worker, as in equation (18). The $M_{it}^{-1/\rho}$ term in (17) is a Benassy

TABLE 1
ECONOMIC ENVIRONMENT

	Equation	Equation Number
Aggregate output	$Y_t = (\int_0^1 Y_{it}^{(\sigma-1)/\sigma} di)^{\sigma/(\sigma-1)}$ with $\sigma > 1$	(16)
Variety i output	$Y_{it} = Z_{Kit}(M_{it}^{-1/\rho} \sum_{j=1}^{M_{it}} x_{ijt}^{(\rho-1)/\rho})^{\rho/(\rho-1)}$ with $\rho > 1$	(17)
Production of ingredients	$x_{ijt} = L_{ijt}$	(18)
Best recipe	$Z_{Kit} = \max_{c=1,\ldots,\tilde{K}_{it}} z_{ic}$, where $\tilde{K}_{it} \sim \text{Poisson}(K_t)$	(19)
Weibull distribution of z_{ic}	$z_{ic} \sim F(x) = 1 - e^{-x^{\beta}}$	(20)
Number of ingredients evaluated	$\dot{N}_t = \alpha R_t^{\lambda} N_t^{\phi}, \phi < 1$	(21)
Number of recipes	$K_t = 2^{N_t}$	(22)
Resource constraint:		
Workers	$L_{it} = \sum_{j=1}^{M_{it}} L_{ijt}$ and $\int_0^1 L_{it} di = L_{yt}$	(23)
R&D	$R_t + L_{yt} = L_t$	(24)
Population growth (exogenous)	$L_t = L_0 e^{g_L t}$	(25)

(1996)-type term that neutralizes the standard love-of-variety effect so that recipes that use more ingredients are neither better nor worse inherently. Instead, the productivity of a recipe is captured completely by its productivity index, z_{ic}.

Let $\tilde{K}_{it}$ denote the number of recipes in the cookbook for variety i at date t. Rather than $\tilde{K}_{it}$ being given directly by K_t, we assume that $\tilde{K}_{it}$ is instead a Poisson random variable with parameter K_t and is independent across varieties.[3] As explained in corollary 1, introducing this bit of randomness has technical value in that the extreme value theorem applies for all K_t rather than asymptotically. Because we have a continuum of varieties with unit measure, the aggregate number of recipes is given by the mean of the Poisson distribution and therefore equals K_t at each point in time.

Each recipe has a productivity that is i.i.d. with $z \sim F(x)$. For now, we assume that the draws are from a Weibull distribution; in section III.C, we will explain how this generalizes.

The Poisson parameter governing the evolution of recipes in the cookbook exhibits combinatorial growth, as defined earlier. That is, $K_t = 2^{N_t}$, where N_t will (eventually) grow at a constant exponential rate. Researchers add ingredients to the cookbook and learn the productivities associated with the new recipes. With R_t as the measure of researchers, $\dot{N}_t = \alpha R_t^{\lambda} N_t^{\phi}$ is the flow of new ingredients whose recipes get evaluated each period, where $\lambda > 0$ and $\phi < 1$, as in Jones (1995). The parameter λ allows for "stepping on toes" effects such as duplication, for example, if $\lambda < 1$. The parameter ϕ allows for intertemporal spillovers: as researchers evaluate more ingredients over time, it can get easier via "standing on shoulders" effects ($\phi > 0$) or possibly harder because of "fishing out" effects ($\phi < 0$).

[3] More precisely, the number of recipes added to the cookbook for variety i between date t and date $t + s$ is a Poisson process with arrival rate $K_{t+s} - K_t$, which, given the additivity of Poisson processes, delivers the statement in the main text.

The remainder of table 1 gives the resource constraints for the economy. In short, the sum of all the workers and the researchers is equal to the total population, of measure L_t. And there is exponential population growth at a constant rate g_L.

Does the idea distribution shift out over time?—The model is built around the assumption that there is a single fixed distribution $\bar{F}(x)$ that determines the productivity of all recipes. At some philosophical level, this is arguably a plausible assumption: the space of past, current, and future technologies is a set of recipes and each technology is associated with some productivity. Let $\bar{F}(x)$ be the distribution of these productivities.

When one asks about a shifting distribution, what one really has in mind is that ideas are discovered in some order: it would have been inconceivable that the smartphone was discovered before telephones, radio, and semiconductors. To see how this is already implicitly incorporated here, recall that new ideas are themselves potential ingredients for future recipes. Then the telephone and semiconductors must be invented before the smartphone even though the underlying distribution of possible technologies does not shift.[4]

C. *Solving the Model*

To keep things simple, we consider the allocation that maximizes Y_t at each point in time with a fixed rule of thumb allocation of people between research and working: $R_t = \bar{s}L_t$.

The symmetry in equations (17) and (18) imply that it is efficient to use the same quantity of each ingredient, so that

$$x_{ijt} = x_{it} = \frac{L_{it}}{M_{it}}.$$

Substituting this into the production function in (17) gives

$$Y_{it} = Z_{Kit} L_{it}. \tag{26}$$

Given a number of workers $L_{yt} = (1 - \bar{s})L_t$, the allocation that maximizes Y_t solves

$$\max_{\{L_{it}\}} Y_t = \left(\int_0^1 (Z_{Kit} L_{it})^{(\sigma-1)/\sigma} di \right)^{\sigma/(\sigma-1)} \tag{27}$$

subject to $\int_0^1 L_{it}\, di = L_{yt}$. The solution to this standard CES problem is given by

[4] Alternatively, it would be easy to incorporate a shifting distribution as in Kortum (1997).

$$Y_t = Z_{Kt}(1 - \bar{s})L_t, \text{ where} \tag{28}$$

$$Z_{Kt} = \left(\int_0^1 Z_{Kit}^{\sigma-1}\, di\right)^{1/(\sigma-1)}. \tag{29}$$

Turning to the research side of the model, we have

$$\frac{\dot{N}_t}{N_t} = \frac{\alpha R_t^{\lambda}}{N_t^{1-\phi}} = \frac{\alpha(\bar{s}L_t)^{\lambda}}{N_t^{1-\phi}}.$$

This stable differential equation implies a constant asymptotic growth rate for N_t. In that case, the ratio on the right-hand side of the equation must be constant, which implies that the numerator and denominator grow at the same rate. Therefore,

$$g_N \equiv \lim_{t\to\infty} \frac{\dot{N}_t}{N_t} = \frac{\lambda g_L}{1-\phi}. \tag{30}$$

Given the combinatorial growth process, we then have

$$g_{\log K} = g_N = \frac{\lambda g_L}{1-\phi},$$

and therefore K_t goes to infinity as a double exponential process.

Combining corollary 1 with the Weibull distribution $\bar{F}(x) = e^{-x^{\beta}}$ gives

$$\begin{aligned} K\bar{F}(Z_{Ki}) &= \varepsilon, \\ Ke^{-Z_{Ki}^{\beta}} &= \varepsilon, \\ \Rightarrow \log K - Z_{Ki}^{\beta} &= \log \varepsilon, \\ \Rightarrow Z_{Ki} &= (\log K - \log \varepsilon)^{1/\beta}, \\ \Rightarrow Z_{Ki} &= (\log K)^{1/\beta}\left(1 - \frac{\log \varepsilon}{\log K}\right)^{1/\beta}, \end{aligned} \tag{31}$$

where $\varepsilon \sim G(\varepsilon)$ and $G(\varepsilon)$ is the normalized exponential distribution from corollary 1, with $0 \leq \varepsilon < K$.

Now we can integrate across the different sectors—and change the variable of integration to ε—to get aggregate productivity. Recall that a fraction e^{-K} of our sectors will not have received any draws from the Poisson process; we assume that their productivity is zero. The remaining $1 - e^{-K}$ receive draws, so their productivity in characterized by equation (31) above. Therefore,

$$\begin{aligned}
Z_{Kt} &= \left(\int_0^1 Z_{Kit}^{\sigma-1}\, di\right)^{1/(\sigma-1)} \\
&= \left[e^{-K}\cdot 0 + (1-e^{-K})(\log K)^{(\sigma-1)/\beta}\int\left(1-\frac{\log\varepsilon}{\log K}\right)^{(\sigma-1)/\beta} dG(\varepsilon)\right]^{1/(\sigma-1)} \\
&= (\log K)^{1/\beta}\underbrace{\left((1-e^{-K})\int\left(1-\frac{\log\varepsilon}{\log K}\right)^{(\sigma-1)/\beta} dG(\varepsilon)\right)^{1/(\sigma-1)}}_{\equiv h(K)} \\
&= (\log K)^{1/\beta}\, h(K),
\end{aligned} \tag{32}$$

where $h(K)$ is a particular moment of the $G(\varepsilon)$ distribution that depends on K. More importantly, notice that $h(K)$ converges to one as K goes to infinity and therefore

$$g_Z \equiv \lim_{K\to\infty}\frac{\dot{Z}_{Kt}}{Z_{Kt}} = \frac{g_{\log K}}{\beta} = \frac{g_N}{\beta}$$

and

$$\boxed{g_y = g_Z = \frac{g_N}{\beta} = \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}.} \tag{33}$$

As was suggested by the basic statistical model, we have a setting where output per person, $y \equiv Y/L$, grows exponentially. Valuable new ideas get increasingly hard to find over time at a rate that depends on β, the parameter governing the thinness of the tail of the Weibull distribution. But combinatorial growth in the number of recipes, driven by population growth in the number of researchers, offsets the thinness of the tail and produces exponential growth in incomes. Interestingly, this formulation simultaneously allows for both "ideas get harder to find" via β and "standing on the shoulders of giants" if $\phi > 0$.

IV. Generalizing to Other Distributions

In sections II–III, we characterized the asymptotic growth rate of Z_K when the underlying distribution was Pareto, exponential, or Weibull. In this section, we explain how these results generalize.

A. Relationship with Extreme Value Theory

The classic results in extreme value theory take the following form: let $a_K > 0$ and b_K be normalizing sequences that depend only on K. If $(Z_K - b_K)/a_K$ converges in distribution, then it converges to one of three types, two of which are the Fréchet and the Gumbel mentioned above. Moreover, this convergence occurs if and only if the tail of the distribution behaves in particular ways. In other words, the theorem requires strong assumptions on the underlying $F(x)$. This featured prominently in Kortum (1997) and is given textbook treatment by Galambos (1978), Johnson, Kotz, and Balakrishnan (1995), Embrechts, Mikosch, and Klüppelberg (1997), de Haan and Ferreira (2006), and Resnick (2008).

Interestingly, the result that $K\bar{F}(Z_K)$ converges in distribution to an exponential, as shown in theorem 1, does not require any such assumptions. In particular, essentially all we assumed is that the distribution function is continuous and invertible. At some level of course, this is not surprising: we are applying the distribution function $\bar{F}(\cdot)$ itself to the max, and this undoes the role played by the distribution in the convergence.

B. A General Condition for Combinatorial Growth

Up to this point, we have shown that the exponential and Weibull distributions lead combinatorial growth in the number of draws to produce exponential growth in the max extreme value. In this section and section IV.C, we explain how this result generalizes. We begin by characterizing the set of distributions such that this is true.

THEOREM 2 (A general condition for combinatorial growth). Consider the growth model of section III but with $z_i \sim F(z)$ as a general continuous and unbounded distribution, where $F(\cdot)$ is monotone and differentiable on its support $[z_0, \infty)$, with $z_0 \geq 0$. Let $\eta(x)$ denote the elasticity of the tail cdf $\bar{F}(x)$; that is, $\eta(x) \equiv -d\log \bar{F}(x)/d\log x$. Then

$$\lim_{t\to\infty} \frac{\dot{Z}_{Kt}}{Z_{Kt}} = \frac{g_N}{\alpha} \tag{34}$$

if and only if

$$\lim_{x\to\infty} \frac{\eta(x)}{x^{\alpha}} = \text{Constant} > 0 \tag{35}$$

for some $\alpha > 0$.

Proof.—See appendix B.

It has long been appreciated that constant exponential growth requires power functions, and this result shows that combinatorial growth is no different. The set of distributions that lead to constant exponential growth in the max when draws are combinatorial is the set for which the elasticity of the tail cdf is asymptotically a power function; that is, the elasticity of the elasticity (the superelasticity?) is itself asymptotically constant.[5]

[5] Klenow and Willis (2016) consider demand functions with this property.

Some remarks and examples are helpful to understand this result. First, consider the Kortum (1997) result, where the upper tail must be equivalent to a Pareto distribution. For Pareto, $\bar{F}(x) = x^{-\alpha}$, so $\eta(x) = \alpha$; the elasticity itself is constant. Combinatorial growth moves the constant elasticity down a log derivative. For example, consider the Weibull distribution with $\bar{F}(x) = e^{-x^{\beta}}$. In this case, it is straightforward to show that $\eta(x) = \beta x^{\beta}$; the exponential distribution is the same with $\beta = 1$.

Another useful example is the standard normal distribution, which has tail cdf $\bar{F}(x) = 1 - (1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-u^2/2} du$. The similarity between the normal and the Weibull with $\beta = 2$ is suggested by the fact that the tail of a normal falls with e^{-x^2} and the tail of a Weibull falls with $e^{-x^{\beta}}$. In fact, $\eta(x)$ behaves like x^2 asymptotically in the normal case, just like the Weibull with $\beta = 2$. Therefore, the max over K draws from a normal rises with $(\log K)^{1/2}$, and combinatorial draws from a normal distribution lead to exponential growth at the rate $g_N/2$.[6]

Next, consider a generalized Weibull distribution with $\bar{F}(x) = x^{\gamma} e^{-x^{\beta}}$. In this case, $\eta(x) = \beta x^{\beta} - \gamma$, which is asymptotically a power function with parameter β once again. Or generalizing a different way, suppose $\bar{F}(x) = e^{-(x^{\beta}+x^{\gamma})}$, where $\beta > \gamma$. It is straightforward to show that the asymptotic power exponent is again just β.

Familiar examples of distributions in this class include the normal, the exponential, the Weibull, the Gumbel, the logistic, and the gamma distributions. Additional less familiar examples are provided in section IV.C.

One final remark about theorem 2 is helpful in putting the result into context. There is nothing essential about the number 2 in the expression $K = 2^N$ for generating the result (though it is of course valuable for the combinatorial interpretation). Instead, for example, we could make the base e itself so that $K_t = e^{e^{nt}}$ and the tail of $\bar{F}$ continues to behave like $e^{-x^{\alpha}}$. Compare this with Kortum (1997), where $K_t = e^{nt}$ and $\bar{F}$ looks like $x^{-\alpha}$. We are making the tail exponentially thinner but marching down this thin tail exponentially faster. It just so happens that many conventional distributions have precisely this kind of thin tail, and combinatorial growth is an intuitive example of this double exponential growth.

C. *Scaling and Growth for Other Distributions*

Section IV.B characterized the class of distributions for which combinatorial growth in draws leads to exponential growth in the extreme value.

[6] For the standard normal distribution, $\eta(x) = xe^{-x^2/2}/\bar{F}(x)$ (where we ignore the $1/\sqrt{2\pi}$ since it does not affect the elasticity). Then $\eta(x)/x^2 = e^{-x^2/2}/(x\bar{F}(x))$, and one use of L'Hôpital's rule verifies that this has a constant limit as $x \to \infty$. (The result uses the fact that $\eta(x) \to \infty$ for the normal.)

We now consider some other distributions and use theorem 1 to characterize the max.

First, consider the lognormal distribution. In that case, $\log x$ has a normal distribution. Using the change of variables method and the normal scaling discussed above, we obtain

$$\frac{\max \log x}{(\log K)^{1/2}} \xrightarrow{p} \text{Constant}$$
$$\Rightarrow \frac{\max x}{\exp(\sqrt{\log K})} \xrightarrow{p} \text{Constant}\,.$$

That is, the max grows with $\exp(\sqrt{\log K})$. If $K = 2^N$ and N itself grows exponentially, then the max grows with $\exp(\sqrt{N})$ and $g_Z = 1/2 \cdot g_N \sqrt{N}$, so the growth rate itself grows exponentially.

This is an important and perhaps slightly surprising finding: not all thin-tailed distributions give rise to exponential growth when draws are combinatoric. When x is drawn from a normal distribution, exponential growth emerges. But when $\log x$ is drawn from a normal distribution, the tails are now too thick: we are drawing proportional increments from the normal and those proportional increments grow exponentially, which delivers faster than exponential growth. This same logic applies to other cases: if we find a distribution for which the max x grows as a power function of $\log K$, then if $\log x$ is drawn from that same distribution, its tail will be too thick and combinatorial growth in K will cause the max to explode.[7]

However, one can calculate the growth rate of K that is required to produce exponential growth in Z_K in the lognormal case. Because the max grows with $\exp(\sqrt{\log K})$, we need $\sqrt{\log K} = gt$ and therefore $\log K = (gt)^2$ or $K_t = \exp{(gt)}^2$: the number of draws grows faster than exponentially but slower than combinatorially.

Our next instructive example features tails that are thinner than the class of exponential-like distributions. Consider the Gompertz distribution, which is commonly used by demographers to model life expectancy. Its distribution function is $F(x) = 1 - \exp(-(e^{\beta x} - 1))$ so that its tail is $\bar{F}(x) = \exp(-(e^{\beta x} - 1))$. In other words, the exponential tail of the distribution itself falls off exponentially as $e^{\beta x}$ rather than as a power function like x^β in the Weibull case. The change of variables method works here: assume that y is exponentially distributed, and let $y = e^{\beta x} - 1$ so that x has a Gompertz distribution. Then

[7] To see another interesting application of this fact, suppose that $\log x$ is drawn from an exponential distribution. Notice that this is equivalent to x being drawn from a Pareto distribution. Exponential growth in K delivers exponential growth in the max, as in Kortum (1997). Therefore, combinatorial draws will lead to explosive growth.

$$\frac{\max y}{\log K} \xrightarrow{p} \text{Constant}$$
$$\Rightarrow \frac{\max e^{\beta x} - 1}{\log K} \xrightarrow{p} \text{Constant}$$
$$\Rightarrow \frac{\max e^{\beta x}}{\log K} \xrightarrow{p} \text{Constant}$$
$$\Rightarrow \frac{\max x}{(1/\beta)\log(\log K)} \xrightarrow{p} \text{Constant.}$$

In this case, the max grows with $\log(\log K)$. Exponential growth in the max requires $\log(\log K)$ to grow exponentially. Even combinatorial expansion is not enough: if $K = 2^N$, the max grows with $\log N$, and exponential growth in N yields arithmetic (linear) growth in the max.

Another distribution that features a double exponential is the Gumbel distribution itself, $F(x) = e^{-e^{-x}}$. However, notice that the Gumbel distribution is tail equivalent to the exponential distribution. That is, for x large, $e^{-e^{-x}} \approx 1 - e^{-x}$, so the Gumbel has an exponential upper tail. For this reason, the max grows directly with $\log K$, just like the exponential.

1. Microfoundations for Romer (1990)

There is a final special case worth considering. One of the interesting findings in Kortum (1997) is that in his setup, there did not exist a stationary distribution from which a constant number of draws each period leads to exponential growth in the max. In other words, in Kortum's environment, there was no microfoundation for the Romer (1990) model, in which a constant population leads to exponential growth. However, this turns out to result from the fact that Kortum restricted his setup to one in which the classic extreme value theorem applies (i.e., that an affine transformation of the max converges in distribution). The alternative approach here can be used to derive just such a microfoundation.

Suppose that y is drawn from a Pareto distribution. Let $y = \log x$, and let us say that x has a log-Pareto distribution (analogous to the lognormal): $F(x) = 1 - 1/(\log x)^\alpha$ and $\bar{F}(x) = 1/(\log x)^\alpha$. We could use the change of variables method to get the scaling immediately, but it is even more instructive to go back to equation (5):

$$K\bar{F}(Z_K) = \varepsilon + o_p(1)$$
$$\Rightarrow \frac{K}{(\log Z_K)^\alpha} = \varepsilon + o_p(1) \qquad (36)$$
$$\Rightarrow \frac{\log Z_K}{K^{1/\alpha}} = \left(\frac{1}{\varepsilon + o_p(1)}\right)^{1/\alpha}.$$

Next, because ε is distributed as exponential with mean 1, $\varepsilon^{-1/\alpha}$ is a Fréchet random variable with parameter α.[8] Using this fact in equation (36) gives

$$\frac{\log Z_K}{K^{1/\alpha}} \overset{a}{\sim} \text{Fréchet}(\alpha), \tag{37}$$

and therefore

$$\log Z_K = K^{1/\alpha}(\tilde{\varepsilon} + o_p(1)), \tag{38}$$

where $\tilde{\varepsilon}$ is a Fréchet random variable with parameter α.

To see the microfoundations for Romer (1990), suppose $\Delta K_t = \beta L$, where L is a constant population. Then $K_t = K_0 + gt$ grows linearly where $g \equiv \beta L$ and—if $\alpha = 1$—log Z_K will grow linearly as well, apart from the shocks, which delivers exponential growth in Z_K.[9] In other words, if our productivity draws are log-Pareto distributed with the Pareto parameter equal to 1, we get a microfoundation for the Romer (1990) model.

It is interesting to contrast this result with Kortum (1997). Kortum found that standard extreme value theory could not provide a microfoundation for Romer (1990). Looking at equation (36), we can see why: to get a stationary distribution, we need to take the natural logarithm of Z_K. This is a nonlinear transformation rather than an affine transformation and therefore does not fit the framework of the standard extreme value theory.

Finally, it is worth noting that the microfoundation of the Romer case leads to several counterfactual predictions. For example, according to equation (37), the log of productivity, not the level, would have a Fréchet distribution and therefore a Pareto upper tail. This implies much more inequality in the firm size distribution than we observe (see Axtell 2001; Luttmer 2010). In addition, if K rises linearly, then the variance of log productivity would increase over time.[10] But even that prediction is more complicated than it first appears: for $\alpha = 1$, neither the mean nor the variance of the Fréchet distribution for $\tilde{\varepsilon}$ exists; the tail of the distribution is

[8] Since ε has an exponential distribution with a mean equal to 1,

$$\begin{aligned} e^{-m} &= \Pr[\varepsilon \geq m] \\ &= \Pr\left[\frac{1}{\varepsilon} \leq \frac{1}{m}\right] \\ &= \Pr\left[\left(\frac{1}{\varepsilon}\right)^{1/\alpha} \leq \left(\frac{1}{m}\right)^{1/\alpha}\right]. \end{aligned}$$

Now let $y \equiv \varepsilon^{-1/\alpha}$ and $x \equiv m^{-1/\alpha}$ so that $m = x^{-\alpha}$. With these substitutions, we have

$$\Pr[y \leq x] = e^{-x^{-\alpha}}.$$

[9] The Fréchet distribution now shocks the growth rate, and for $\alpha = 1$, the tail of the Fréchet distribution is so thick that the mean of these shocks does not exist.

[10] For this to hold, suppose $\alpha > 2$, so the variance of the Fréchet distribution exists.

too thick. All of this is to say that I see the microfoundations for the Romer case as an interesting illustration of the technique, not as providing a realistic model of growth.

2. Summary

These results are collected together in table 2. In particular, they show how the max Z_K scales as a function of K.

In Kortum (1997), an exponentially growing number of draws from any distribution in the Fréchet domain of attraction leads to exponential growth in the max. One might have conjectured that combinatorial growth would work the same way. In particular, a natural guess is that all distributions in the basin of attraction of the Gumbel distribution could deliver exponential growth in productivity when the number of draws grows combinatorially. This guess turns out to be wrong. The set of distributions in the Gumbel basin of attraction is large and includes slightly thick tails like the lognormal; thin tails like the normal, exponential, gamma, and the Gumbel itself; as well as even thinner tails, like the Gompertz. Only the intermediate class delivers exponential growth in the max for combinatorially growing draws.

V. Evidence

One of the facts that Kortum (1997) sought to explain was the time series of patents in the United States. In particular, Kortum emphasized the relative stability of patents: the number of patents granted to US inventors in 1915, 1950, and 1985 was roughly the same, around 40,000. In his setup, each new patent is on average a proportional improvement on

TABLE 2
Scaling of Z_K for Various Distributions

Distribution	cdf (1)	b_K (2)	$b_K(N)$ for $K = 2^N$ (3)	Growth Rate for $K = 2^N$ (4)
Exponential	$1 - e^{-\theta x}$	$\log K$	N	g_N
Gumbel	$e^{-e^{-x}}$	$\log K$	N	g_N
Weibull	$1 - e^{-x^\beta}$	$(\log K)^{1/\beta}$	$N^{1/\beta}$	$\frac{g_N}{\beta}$
Normal	$\frac{1}{\sqrt{2\pi}}\int e^{-x^2/2}dx$	$(\log K)^{1/2}$	$\sqrt{N}$	$\frac{g_N}{2}$
Lognormal	$\frac{1}{\sqrt{2\pi}}\int e^{-(\log x)^2/2}dx$	$\exp(\sqrt{\log K})$	$e^{\sqrt{N}}$	$\frac{g_N}{2}\cdot\sqrt{N}$
Gompertz	$1 - \exp(-(e^{\beta x} - 1))$	$\frac{1}{\beta}\log(\log K)$	$\frac{1}{\beta}\log N$	Arithmetic
Log-Pareto	$1 - \frac{1}{(\log x)^\alpha}$	$\exp(K^{1/\alpha})$	...	...

Note.—In all rows except the final one, $Z_K/b_K \xrightarrow{p}$ Constant. The final row is more subtle, as discussed in the main text. Columns 3 and 4 focus on the combinatorial case. Column 3 translates this into scaling with N for $K = 2^N$ (ignoring some multiplicative constants). Column 4 shows the asymptotic growth rate of Z_K if $N(t)$ grows exponentially at rate g_N.

the previous state of the art, so a constant flow of new patents generates stable exponential growth.

To see this point, we first have to define what we mean by a patent or a valuable new idea in the model. We follow Kortum (1997) in defining the flow of new patents as the flow of new ideas that are improvements over the state of the art. Recall the intuition we developed for theorem 1 back in equation (6). In particular, the probability that a new idea exceeds the current frontier is $1/K$ on average:

$$\bar{F}(Z_K) = \frac{1}{K}(\varepsilon + o_p(1)). \tag{39}$$

What does this imply about the flow of patents in the growth model? With $\dot{K}_t$ new ideas being discovered at date t and the fraction $1/K_t$ exceeding the frontier, the time series of patents in the model is on average $\dot{K}_t/K_t$. This is precisely the logic in Kortum (1997), and it is therefore easy to see how the flow of patents could be constant in that setup. With the Pareto distribution, ideas are large—proportional improvements—so a constant flow of new patents can generate exponential growth.

In the combinatorial model, however, this quantity is not constant. Instead, first consider the model in which $\dot{N}_t = \alpha R_t$ (i.e., $\lambda = 1$ and $\phi = 0$):

$$\begin{aligned} K_t &= 2^{N_t} \\ \Rightarrow \frac{\dot{K}_t}{K_t} &= \log 2 \cdot \dot{N}_t \\ &= \log 2 \cdot \alpha R_t \\ &= \log 2 \cdot \alpha \bar{s} L_0 e^{g_L t}. \end{aligned} \tag{40}$$

That is, the number of new patents in the combinatorial model grows exponentially over time. In fact, the number of patents per researcher would actually be constant in this case. More generally, if one allows for $\lambda \neq 1$ or $\phi \neq 0$, the number of patents will (asymptotically) exhibit exponential growth and the number of patents per researcher can either decline or increase over time.[11]

The intuition for this result is straightforward: because of the thin tail of the probability distribution, the typical new patent is only slightly better than the previous state of the art: ideas are small on average. Exponential growth in productivity requires us to march down the tail very quickly. Because new ideas are small, we need an exponentially growing number

[11] Kogan et al. (2017) document that patents per capita were relatively stationary between 1930 and 1990 but have risen since then. The pre-1990 evidence would be consistent with the combinatorial model with $\phi = 0$, while the period since 1990 is more consistent with $\phi > 0$.

of them to increase productivity proportionally, and this corresponds to exponential growth in the flow of patents in the model.

In summary, the Kortum (1997) setup implies that the flow of new patents will be constant. The combinatorial model instead implies that the flow of new patents will rise exponentially. What does the evidence say?

Figure 1 shows the time series for patents granted by the US Patent Office both in total (i.e., including foreign inventors) and to US inventors only. Far from being constant, the patent series viewed from the perspective of today looks much more like a series that itself exhibits exponential growth. This is especially true for the "Total" series, which is surely the most relevant: growth in a country depends on ideas that are used there, regardless of where they are invented. Put differently, in the Kortum (1997) setup, the rise in patents in the United States would imply an eightfold increase in the rate of economic growth, something we certainly do not see in the data.

One resolution is that perhaps the meaning of a patent has changed over time. Legal reforms and other changes may imply that a patent in 2020 is not the same as a patent in 1980 (Kortum and Lerner 1998; Hall and Ziedonis 2001); if they are not comparable, then one cannot view this graph as telling us about the behavior of ideas over time. Perhaps a true series for new ideas is actually constant.

Kelly et al. (2021) use machine learning methods on the text of patents to measure breakthrough patents, defined as those that are novel relative

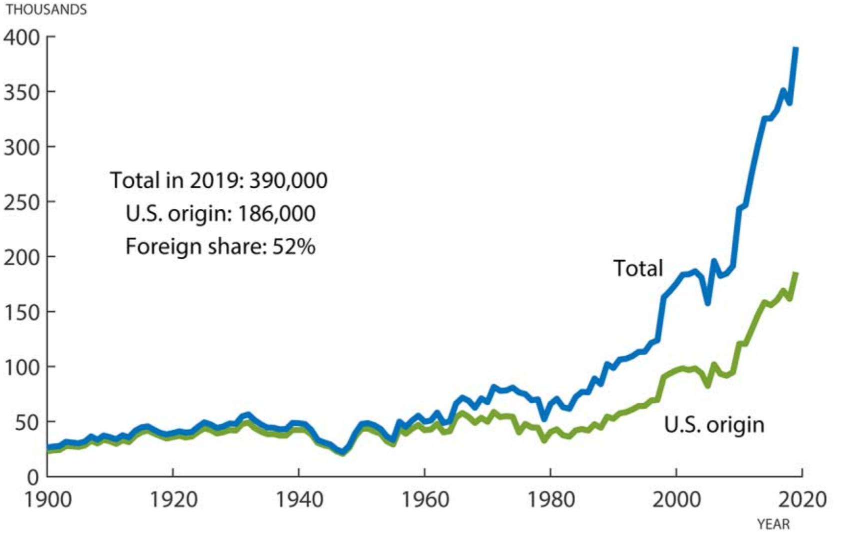

Fig. 1.—Patents granted by US Patent and Trademark Office. Source: US Patent and Trademark Office (2020).

to previous patents and that are related to subsequent innovations. One might suppose that breakthrough ideas would be patented under both the pre-1980 and post-1980 regimes, so this is one way to address these concerns. Their aggregate time series for breakthrough patents is shown in figure 2. Two things stand out. First, the time series is far from constant. For example, between 1900 and 1980, the number of breakthrough patents rises by a factor of 4, and over the entire century, the flow of breakthrough patents increases by more than a factor of 20. That is, breakthrough patents look very different from the constant that is implied by Kortum (1997). On the other hand, the time series does not exhibit stable exponential growth. Instead, it rises steadily between 1900 and the end of the 1930s before plummeting during World War II. It is then relatively stable between 1950 and 1980 before rising sharply after 1980 (by something like a factor of 8 between 1980 and 2002). Long and variable lags and measurement issues are surely relevant; otherwise, it is hard to understand the stability of gross domestic product per-person growth rates during the twentieth century.

Another important factor is of course the rise of computers and information technology. For example, the growth rate of breakthrough patents from figure 2 between 1950 and 2002 is 3.9% per year. In part, this reflects an extraordinary growth rate in computers/electronics of 9.1% per year. Nevertheless, even omitting computers/electronics entirely, the growth rate of the remaining breakthrough patents still averages 2.0% per year since 1950.

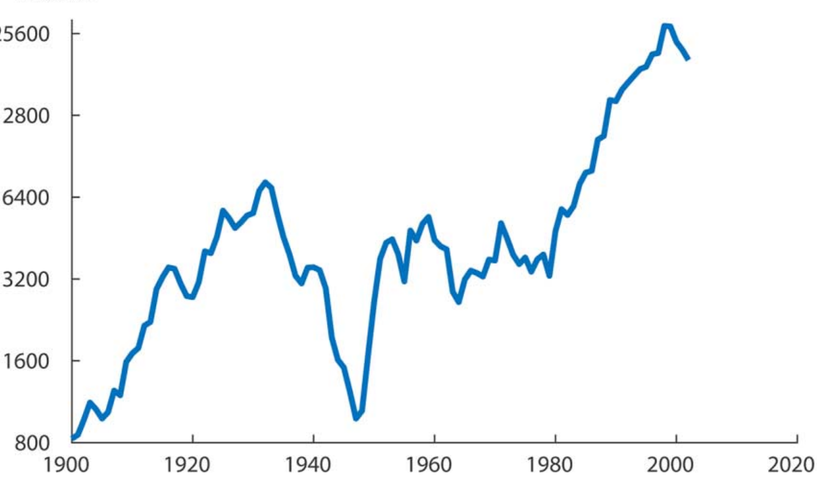

Fig. 2.—Breakthrough patents from Kelly et al. (2021). Source: author's calculations using data from Kelly et al. (2021).

Figure 3 shows related evidence by plotting the average annual growth rate of patents granted by the US Patent and Trademark Office for 129 technology classes over the period 1950–90, that is, before the explosion of patenting associated with legal changes. Only eight of the technology classes show declines in patenting over this period, and this is primarily in classes related to industries that are either in decline or offshored, such as leather/pelts (C14), railways (B61), and textile treatments (D06). The other 121 classes show positive and typically substantial rates of growth in patenting; the weighted average of the growth rates is 3.6% per year. Including more recent data (not shown) would only reinforce this point: between 1950 and 2019, only a single technology class (leather/pelts [C14]) displays a decline in patenting.

Alternatively, we can also consider a different measure of innovation: academic publications. Figure 4 shows that exponential growth also characterizes annual publication counts. Depending on the measure used, publications grew at between 3.3% and 4.4% per year, increasing overall by a factor of between 5 and 9 since 1970.

While none of these measures is perfect—and indeed, one drawback of the innovation literature is that we do not have solid measures of

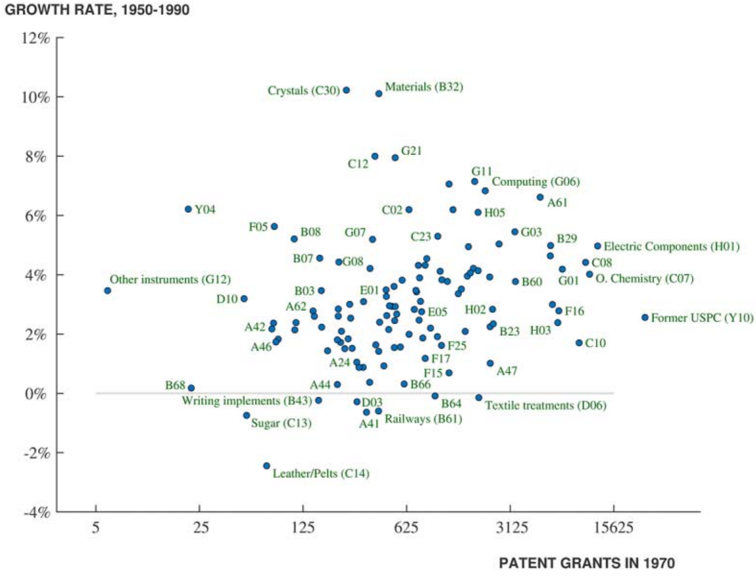

FIG. 3.—US patent growth by technology class, 1950–90. The vertical axis shows the average annual growth rate of patents granted by the US Patent and Trademark Office for 129 technology classes. The horizontal axis shows the corresponding number of patents granted in the year 1970. Source: author's calculation using data provided by Amit Seru.

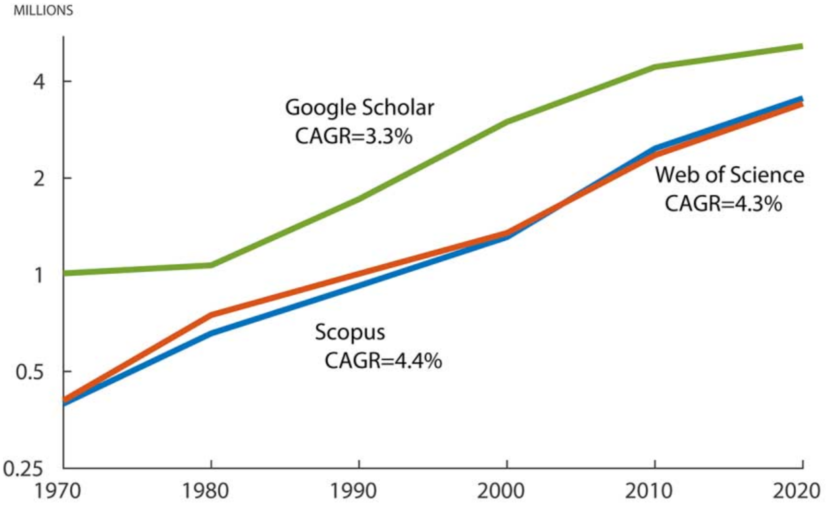

FIG. 4.—Annual academic publication counts, 1970–2020. CAGR = cumulative average annual growth rate over the period. Source: author's calculations using Google Scholar (2021), Scopus (2021), and Web of Science (2021).

innovation—they all suggest that valuable new ideas may well be characterized as growing over time rather than being constant. Though more work and better idea measures are surely needed, the basic evidence appears to be more consistent with the combinatorial model with thin-tailed distributions rather than the Kortum (1997) model.

Can researchers evaluate a combinatorially growing number of recipes?—This is now a good place to discuss one of the features of the model that might raise a question. An implication of our setup is that researchers are evaluating the productivity of a rapidly increasing number of recipes over time: they each evaluate the recipes associated with, say, α new ingredients each period, but the number of recipes that can be formed from the new and existing number of ingredients grows combinatorially. Is it possible for researchers to evaluate a combinatorially growing number of recipes to find the best one?

We have several responses to this question. The first is the empirical evidence provided above: the combinatorial process leads to exponential growth in valuable new ideas, which is a good description of the data itself. Second, and more philosophically, perhaps it is only the truly good ideas that take time to evaluate: Akerlof's chicken ice cream can be discarded quickly. Chess grandmasters sort through combinatorially growing lines of play with remarkable speed and often find the best move according to computers that search billions of moves per second (Sadler and Regan 2019). The number of truly new ideas grows exponentially precisely with the number of researchers in equation (40), so each researcher would need to devote time to a constant number of new ideas, which seems reasonable.[12] Finally, Jones (2009) and Pearce (2022) show that the size of research teams has been rising over time, and Akcigit, Kerr, and Pearce (2022) show that the number of technologies combined in each patent has also been rising over time. Perhaps these two trends are related: it requires larger teams to successfully evaluate the productivity of the best recipes. In the context of the model here, this would be consistent with $\phi < 0$.

VI. Discussion and Further Connections to the Literature

This section explores various implications of the setup and connections to the literature.

A. *Markets, Markups, and the Cross Section Distribution of Productivity*

This paper takes a shortcut in considering the allocation of resources by focusing on an allocation with a constant fraction of labor engaged in research and with the allocation of the remaining workers across varieties set to maximize output. It would definitely be valuable to study the market allocation as well as the equilibrium distribution of markups and productivities across varieties. Let me make a few observations about why that is even more interesting than it first appears and a valuable direction for future research.

First, recall that in Kortum (1997), the distribution of productivity across varieties is Fréchet. This is hinted at in equation (7). With thin-tailed distributions like the exponential or normal, the distribution of the max is asymptotically Gumbel. An example of this is provided in equation (9), repeated here:

[12] The mathematician Henri Poincaré ([1910] 2000, 88) advocates this view: "To invent, I have said, is to choose; but the word is perhaps not wholly exact. It makes one think of a purchaser before whom are displayed a large number of samples, and who examines them, one after the other, to make a choice. Here the samples would be so numerous that a whole lifetime would not suffice to examine them. This is not the actual state of things. The sterile combinations do not even present themselves to the mind of the inventor. Never in the field of his consciousness do combinations appear that are not really useful, except some that he rejects but which have to some extent the characteristics of useful combinations. All goes on as if the inventor were an examiner for the second degree who would only have to question the candidates who had passed a previous examination." I thank Matt Clancy for this reference.

$$\frac{Z_K}{\log K} = \frac{1}{\theta}\left(1 - \frac{\log(\varepsilon + o_p(1))}{\log K}\right).$$

Recall that ε is an exponentially distributed random variable, so $\log \varepsilon$ has a Gumbel distribution. In other words, the cross section distribution of productivity in this setup is Gumbel instead of Fréchet. However, a close look at the equation reveals that this is not the end of the story. In particular, the Gumbel random variable is divided by $\log K$, so that over time, the variance of detrended log productivity shrinks to zero. In other words, while the Kortum (1997) setup provides microfoundations for heterogeneity in productivity, a model with draws from a thin-tailed distribution does not, at least not in the long run. Evidently, if the combinatorial framework is correct, one needs a different theory of heterogeneity in productivity across firms and varieties.

This same point repeats in other interesting ways. For example, now consider the distribution of markups. When ideas are drawn from a Pareto distribution, Kortum (1997) shows that ideas are "large." As discussed above, a constant flow of patents can generate exponential growth in productivity. This is because each new patent is, on average, a proportional improvement over the preceding max. This proportionality gives rise to a distribution of markups in which the average markup is also governed by the parameter of the Pareto distribution.

In contrast, when ideas are drawn from a thin-tailed distribution, as in the combinatorial setup, we showed above that ideas are "small." That is, it requires exponential growth in the flow of patents to generate exponential economic growth. Because ideas are small improvements, the gap between ideas is small, and therefore so are markups. For example, it is well known that for the exponential distribution with parameter θ, the expected gap between the maximum and the second-largest value is $1/\theta$. That is, additive gaps are constant, so proportional gaps are shrinking at a rate that depends on $\log K$.[13]

In other words, the Pareto distributions associated with Kortum (1997) and a large literature that builds on that paper provides a theory of heterogeneous productivity and stable but heterogeneous markups. However, the combinatorial approach with thin-tailed distributions does not: the distribution of productivity and markups would shrink over time. This can be read as a shortcoming, but instead I think it is an opportunity. Perhaps, motivated by the data on patents and ideas, the combinatorial approach is correct. This means that we need a new and richer theory of heterogeneous markups and productivity, because the idea distribution no longer provides it.

[13] For example, see Arnold, Balakrishnan, and Nagaraja (2008, 73), eq. (4.6.6).

Finally, this brings us back to the market equilibrium. If markups were determined by the gap between the most productive and second most productive firm, proportional markups would fall to zero, causing the return to innovation shrink. Exponential growth in such an equilibrium would decline rather than being constant because of the declining markups. This is another reason why having a richer equilibrium that delivers stable markups is important. I verify (see app. C) that the optimal allocation features sustained exponential growth, so the problem is not with the economic environment. Instead, the standard approach to setting up an equilibrium seems destined to fail in that it would produce an allocation that is far from optimal and does not match the facts about markups and productivity. A richer theory of markups and creative destruction in thin-tailed environments is a logical next step for future research.

B. Correlation

What if the draws from the search distribution $\bar{F}(x)$ are correlated for recipes that share many ingredients? This would be a useful extension to explore but is beyond the scope of the present paper. Most of the results in the extreme value literature, for example, consider the independently and identically distributed (i.i.d.) case. Still, broader results are possible. For example, if the correlation dies off quickly, there are results related to blocks of draws that can be viewed as i.i.d. In this sense, the result is likely to generalize to cases with correlation.

C. Other Applications

Many papers in different literatures build on Kortum (1997) and assume a distribution with a Pareto upper tail. Beyond economic growth, examples include the extensive literature on international trade following Eaton and Kortum (2002) and models of heterogeneous productivity based on Hopenhayn (1992) and Melitz (2003). In each of these cases, it may be productive to consider the combinatorial approach with thin-tailed distributions developed here.

Two more specific applications are worth noting. First are the labor search models of Barlevy (2008) and Martellini and Menzio (2020). The latter studies a search and matching model of the labor market, seeking to understand why technological progress in matching has not led to trends in unemployment or vacancy rates. They show that if match qualities are drawn from a Pareto distribution, then improvements in search technologies—which would tend to increase the frequency of matching—lead to perfectly offsetting effects that leave unemployment and vacancy rates unchanged. In particular, better search technologies also raise workers' reservation quality because it is easier for workers to find new matches.

How would this work with combinatorial matching and thin-tailed distributions?

Finally, a very interesting direction for future research is related to Lucas and Moll (2014), Perla and Tonetti (2014), and the extensive literature that has built on these papers. The basic insight in these papers is related to Kortum (1997): an exponentially growing number of draws (e.g., because of meetings between firms or people) from a Pareto distribution can generate exponential growth and an evolving distribution of heterogeneous productivities. Can combinatorics and thin-tailed distributions replace the continuum and Pareto assumptions underlying these papers?

D. Conclusion

In the end, the paper can be read in two ways. First, there is the "Weitzman meets Kortum" interpretation: if we have the number of draws grow combinatorially then we do not need thick-tailed Pareto distributions to generate economic growth. Instead, draws from standard distributions with thin exponential tails are sufficient. Second, there is a broader contribution embodied in theorem 1. When we consider the max Z_K over K i.i.d. draws from a distribution with tail distribution function $\bar{F}(x)$, the transformed random variable $K\bar{F}(Z_K)$ asymptotically has an exponential distribution under very weak conditions. This result can be used to characterize the way in which the max Z_K increases for any continuous distribution $\bar{F}(x)$ and any time path of (large) K.

Data Availability

Code replicating the figures in this article can be found in Jones (2023) in the Harvard Dataverse, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/X0YWXF.

Appendix

A1. Proof of Corollary 1

Proof. Let $M_p \equiv K\bar{F}(Z_K)$ denote a new random variable, conditional on $P = p$. Given that Z_K is the max over P i.i.d. draws, exactly the same steps used in proving theorem 1 give

$$\Pr\left[M_p \geq m\right] = \left(1 - \frac{m}{K}\right)^p$$

when $p > 0$.

Now we use the Poisson assumption to get the unconditional distribution. Importantly, notice that it is only when the realized number of draws P is greater than zero that the problem is well defined; if there are zero draws to consider, there is nothing to take the max over. Recall that $\Pr[P = p] = e^{-K}K^p/p!$ so that $\Pr[P = 0] = e^{-K}$ and $\Pr[P > 0] = 1 - e^{-K}$. Therefore, for $0 \le m < K$,

$$\begin{aligned}
\Pr[K\bar{F}(Z_K) \ge m] &= \sum_{p=1}^{\infty} \Pr\left[M_p \ge m\right] \cdot \Pr[P = p \mid P > 0] \\
&= \sum_{p=1}^{\infty} \left(1 - \frac{m}{K}\right)^p \cdot \frac{\Pr[P = p]}{\Pr[P > 0]} \\
&= \frac{1}{\Pr[P > 0]} \sum_{p=1}^{\infty} \left(1 - \frac{m}{K}\right)^p \cdot \frac{e^{-K}K^p}{p!} \\
&= \frac{1}{\Pr[P > 0]} \left[\sum_{p=1}^{\infty} \left(1 - \frac{m}{K}\right)^p \cdot \frac{e^{-K}K^p}{p!} + e^{-K} - e^{-K}\right] \\
&= \frac{1}{\Pr[P > 0]} \left[\sum_{p=0}^{\infty} \left(1 - \frac{m}{K}\right)^p \cdot \frac{e^{-K}K^p}{p!} - e^{-K}\right] \\
&= \frac{1}{\Pr[P > 0]} \left[e^{-m} \sum_{p=0}^{\infty} \frac{e^{-K(1-m/K)}\left(K(1 - m/K)\right)^p}{p!} - e^{-K}\right] \\
&= \frac{e^{-m} - e^{-K}}{1 - e^{-K}},
\end{aligned}$$

where the last step uses the fact that the summation term is just the probability that any number of events occurs for a Poisson distribution with parameter $K(1 - m/K)$, that is, the value of the cdf at infinity, which is equal to 1. QED

A2. *Proof of Theorem 2*

Here we prove theorem 2, which provides a necessary and sufficient condition on the shape of the search distribution for combinatorial growth in the draws to deliver exponential growth in the max extreme value.

A2.1. Lemma 1 and Its Proof

In proving this result, the following lemma is very helpful, as it allows us to go back and forth between the elasticity of $\bar{F}$ and the elasticity of $\bar{F}^{-1}$. We will use the notation $\sim$ to denote the following type of convergence: $f(x) \sim x^{\alpha}$ is equivalent to $\lim_{x\to\infty} f(x)/x^{\alpha} = \text{Constant}$.

LEMMA 1. Let $y = \bar{F}(x)$, where $\bar{F}$ is a continuous, differentiable, and invertible function. Then

$$-\frac{d \log \bar{F}(x)}{d \log x} \sim x^{\alpha}$$

if and only if

$$-\frac{d \log \bar{F}^{-1}(y)}{d \log y} \sim [\bar{F}^{-1}(y)]^{-\alpha}$$

(recognizing that the relevant limits are as $x \to \infty$ and therefore $y = \bar{F}(x) \to 0$).

Proof. Let $h(y) \equiv \bar{F}^{-1}(y)$. Applying the function $\bar{F}$ to both sides gives

$$\begin{aligned} y &= \bar{F}(h(y)), \\ \log y &= \log \bar{F}(h(y)), \\ d\log y &= \frac{d\log \bar{F}(h(y))}{d\log h(y)} \cdot d\log h(y). \end{aligned}$$

Rearranging then gives

$$\frac{d\log h(y)}{d\log y} = \left[\frac{d\log \bar{F}(h(y))}{d\log h(y)}\right]^{-1}$$

and therefore

$$\frac{d\log \bar{F}^{-1}(y)}{d\log y} = \left[\frac{d\log \bar{F}(h(y))}{d\log h(y)}\right]^{-1}.$$

Then the result is obvious. If $-d\log \bar{F}(x)/d\log x \sim x^{\alpha}$, then $-d\log \bar{F}^{-1}(y)/d\log y \sim [\bar{F}^{-1}(y)]^{-\alpha}$ and vice versa since $y = \bar{F}(x)$. QED

A2.2. Proof of Theorem 2

We are now ready to prove theorem 2.

Proof. By corollary 1, we have

$$K_t \bar{F}(Z_{Kit}) = \varepsilon,$$

where $\varepsilon \sim G(\varepsilon)$ and $G(\varepsilon)$ is the normalized exponential distribution from corollary 1, with $0 \le \varepsilon < K$.

Inverting the distribution function and solving for Z_{Kit} gives

$$Z_{Kit} = \bar{F}^{-1}\left(\frac{\varepsilon}{K_t}\right).$$

Recall the definition of aggregate productivity: Z_{Kt} is a power mean of the individual variety productivities. Changing the variable of integration from i to ε to take advantage of the continuum of varieties and recalling that the fraction e^{-K_t} of sectors has zero Poisson draws and therefore zero productivity gives

$$\begin{aligned} Z_{Kt}^{\sigma-1} &= (1 - e^{-K_t}) \int Z_{K\varepsilon t}^{\sigma-1}\, dG(\varepsilon) \\ &= (1 - e^{-K_t}) \int \left[\bar{F}^{-1}\left(\frac{\varepsilon}{K_t}\right)\right]^{\sigma-1} dG(\varepsilon). \end{aligned}$$

To simplify the notation, define $h(y) = \bar{F}^{-1}(y)$. Taking logs and differentiating both sides of the above equation with respect to time gives

$$
\begin{aligned}
(\sigma-1)\frac{\dot{Z}_{Kt}}{Z_{Kt}} &= \frac{e^{-K_t}}{1-e^{-K_t}}\frac{dK_t}{dt} + \frac{\sigma-1}{Z_{Kt}^{\sigma-1}}\int\left[h\left(\frac{\varepsilon}{K_t}\right)\right]^{\sigma-2} h'\left(\frac{\varepsilon}{K_t}\right)\left(-\frac{\varepsilon}{K_t^2}\right)\frac{dK_t}{dt}\,dG(\varepsilon) \\
&= \frac{e^{-K_t}}{1-e^{-K_t}}\frac{dK_t}{dt} + \frac{\sigma-1}{Z_{Kt}^{\sigma-1}}\int\left[h\left(\frac{\varepsilon}{K_t}\right)\right]^{\sigma-1}\left(-\frac{h'(\varepsilon/K_t)\cdot\varepsilon/K_t}{h(\varepsilon/K_t)}\right)\frac{\dot{K}_t}{K_t}\,dG(\varepsilon) \\
&= \frac{e^{-K_t}}{1-e^{-K_t}}\frac{dK_t}{dt} + \frac{\sigma-1}{Z_{Kt}^{\sigma-1}}\int\left[\bar{F}^{-1}\left(\frac{\varepsilon}{K_t}\right)\right]^{\sigma-1}\left(-\frac{d\log\bar{F}^{-1}(\varepsilon/K_t)}{d\log(\varepsilon/K_t)}\right)\frac{\dot{K}_t}{K_t}\,dG(\varepsilon).
\end{aligned}
$$

Rearranging the terms slightly and taking limits gives

$$
\lim_{t\to\infty}\frac{\dot{Z}_{Kt}}{Z_{Kt}} = \int \lim \frac{h(\varepsilon/K_t)^{\sigma-1}}{\int h(\varepsilon/K_t)^{\sigma-1}\,dG(\varepsilon)}\cdot\lim\left(-\frac{d\log\bar{F}^{-1}(\varepsilon/K_t)}{d\log(\varepsilon/K_t)}\right)\frac{\dot{K}_t}{K_t}\,dG(\varepsilon), \tag{41}
$$

where we have used the fact that e^{-K_t} goes to zero to eliminate the first term.

A2.2.1. Only If. At this point, we are ready to consider the two directions of the proof. We begin with the "only if" portion. In particular, we can apply lemma 1 to see that $-(d\log\bar{F}^{-1}(\varepsilon/K))/(d\log(\varepsilon/K_t)) \sim \bar{F}^{-1}(\varepsilon/K_t)^{-\alpha}$, which gives

$$
\lim_{t\to\infty}\frac{\dot{Z}_{Kt}}{Z_{Kt}} = \int \lim \frac{h(\varepsilon/K_t)^{\sigma-1}}{\int h(\varepsilon/K_t)^{\sigma-1}\,dG(\varepsilon)}\cdot\lim\frac{\psi\dot{K}_t/K_t}{\bar{F}^{-1}(\varepsilon/K_t)^{\alpha}}\,dG(\varepsilon), \tag{42}
$$

where ψ is the limiting factor of proportionality from the elasticity term.

Now consider the limit of the second key term in equation (42) for each fixed value of ε and use the combinatoric growth of K_t:

$$
\begin{aligned}
v_t &\equiv \frac{\psi\dot{K}_t/K_t}{\bar{F}^{-1}(\varepsilon/K_t)^{\alpha}} \\
&= \frac{\psi\dot{N}_t\log 2}{\bar{F}^{-1}(\varepsilon/K_t)^{\alpha}} \\
&= \text{Constant}\frac{\psi e^{g_N t}}{\bar{F}^{-1}(\varepsilon/K_t)^{\alpha}},
\end{aligned}
$$

where the last expression uses the fact that N_t grows at a constant exponential rate.[14]

By inspection, the limit of v_t is ∞/∞ as $t\to\infty$, so we apply L'Hôpital's rule to get the limit:

$$
\begin{aligned}
\lim v_t &= \lim \text{Constant}\frac{\psi g_N e^{g_N t}}{\alpha\bar{F}^{-1}(\varepsilon/K_t)^{\alpha-1}(\bar{F}^{-1})'(\varepsilon/K_t)(-\varepsilon/K_t^2)\dot{K}_t} \\
&= \frac{g_N}{\alpha}\cdot\lim\frac{\text{Constant}e^{g_N t}}{\dot{K}_t/K_t}\cdot\lim\frac{\psi}{[\bar{F}^{-1}(\varepsilon/K_t)]^{\alpha}\cdot\left(-\left(d\log\bar{F}^{-1}(\varepsilon/K_t)\right)/(d\log(\varepsilon/K_t))\right)} \\
&= \frac{g_N}{\alpha},
\end{aligned}
$$

where the last two terms in the penultimate equation each are equal to 1.

[14] This is easiest in the case where $N_t = N_0 e^{g_N t}$ is just assumed but also holds exactly for $\dot{N} = \alpha R_t = \alpha\bar{s}L_t$ when $\lambda = 1$ and $\phi = 0$ or asymptotically when $\lambda > 0$ and $\phi < 1$.

Finally, substituting this expression in for the limit of v_t back into equation (42) gives

$$\lim_{t\to\infty}\frac{\dot{Z}_{Kt}}{Z_{Kt}} = \frac{g_N}{\alpha}\lim\int\frac{h(\varepsilon/K_t)^{\sigma-1}}{\int h(\varepsilon/K_t)^{\sigma-1}dG(\varepsilon)}\,dG(\varepsilon)$$

$$= \frac{g_N}{\alpha}.$$

That completes the "only if" part of the proof.

A2.2.2. If. Now return to equation (41) for the "if" direction: if $\lim(\dot{Z}_{Kt}/Z_{Kt}) = g_N/\alpha$, then $\eta(x)$ is asymptotically a power function with exponent α. Applying this condition to (41) gives

$$\frac{g_N}{\alpha} = \int\lim\frac{h(\varepsilon/K_t)^{\sigma-1}}{\int h(\varepsilon/K_t)^{\sigma-1}dG(\varepsilon)}\cdot\lim\left(-\frac{d\log\bar{F}^{-1}(\varepsilon/K_t)}{d\log(\varepsilon/K_t)}\right)\frac{\dot{K}_t}{K_t}\,dG(\varepsilon).$$

The first term on the right-hand side of this expression is a collection of weights that integrate to a value of 1 for all K_t. Therefore, this term does not trend over time. Since the left-hand side is constant, though, this means that the second term on the right-hand side must also be constant. In particular, this means that the elasticity term must decline exponentially at the rate g_N. Defining $v(K)$ to be this elasticity, we have

$$v(K) \equiv -\frac{d\log\bar{F}^{-1}(\varepsilon/K_t)}{d\log(\varepsilon/K_t)},$$

and we require

$$v(K)\frac{\dot{K}_t}{K_t}\to\frac{g_N}{\alpha}.$$

Now recall $K = 2^N$ so that $\dot{K}t/Kt = \dot{N}\log 2$ and therefore

$$\frac{\frac{\dot{K}_t}{K_t}}{\alpha\log K} = \frac{\dot{N}_t\log 2}{\alpha N_t\log 2}$$

$$\to\frac{g_N}{\alpha}.$$

Combining these last two expressions means that we require

$$v(K)\alpha\log K\to 1.$$

Let $y\equiv\varepsilon/K$ for a fixed ε. Substituting into the previous expression gives

$$\left[-\frac{d\log\bar{F}^{-1}(y)}{d\log y}\right][-\alpha\log y]\to 1$$

since $-\log y/\log K\to 1$ for a fixed ε.

To finish the proof, we write this equation in terms of $-\log y$, which is positive since $0 < y < 1$. We also switch to the ~ version of this equation (being sure to keep α since the convergence is to 1 rather than to any constant) and then integrate:

$$\begin{aligned}
\frac{d\log \bar{F}^{-1}(y)}{d(-\log y)} &\sim \frac{1}{\alpha}\cdot\frac{1}{(-\log y)}\\
\Rightarrow d\log \bar{F}^{-1}(y) &\sim \frac{1}{\alpha}\cdot\frac{d(-\log y)}{(-\log y)}\\
\Rightarrow \int d\log \bar{F}^{-1}(y) &\sim \frac{1}{\alpha}\cdot\int\frac{d(-\log y)}{(-\log y)}\\
\Rightarrow \log \bar{F}^{-1}(y) &\sim \text{Constant} + \frac{1}{\alpha}\log(-\log y)\\
\Rightarrow \bar{F}^{-1}(y) &\sim \text{Constant}\left[e^{\log(-\log y)}\right]^{1/\alpha}\\
\Rightarrow x &\sim (-\log y)^{1/\alpha}\\
\Rightarrow -\log y &\sim x^{\alpha}\\
\Rightarrow -\log \bar{F}(x) &\sim x^{\alpha}\\
\Rightarrow -\frac{d\log \bar{F}(x)}{dx} &\sim \alpha x^{\alpha-1}\\
\Rightarrow -\frac{d\log \bar{F}(x)}{d\log x} &\sim x^{\alpha},
\end{aligned}$$

where we use the notation $y = \bar{F}(x)$ and take advantage of the ~ notation to drop the (positive) constants whenever convenient. QED

A3. The Optimal Allocation

In this section, we characterize the optimal allocation for the economic environment in table 1 and show that it features an asymptotic balanced growth path with an interior solution for the fraction of labor devoted to R&D. To make this a well-defined problem, we need to add a bit more structure: a standard utility function with rate of time preference θ, flow utility $u(c) = c^{1-\gamma}/(1-\gamma)$, and the resource constraint that $c_t = Y_t/L_t$. As explained in section III.C, the symmetry of the setup and the fact that labor will be allocated across varieties to maximize output means that the optimal allocation involves choosing the time path of research intensity, s_t:

$$\begin{aligned}
&\max_{\{s_t\}} \int_0^{\infty} e^{-\theta t}u(c_t)\,dt \text{ subject to}\\
c_t &= Y_t/L_t = Z_{Kt}(1-s_t)\\
Z_{Kt} &= (\log K_t)^{1/\beta}h(K_t)\\
K_t &= 2^{N_t}\\
\dot{N}_t &= \alpha R_t^{\lambda}N_t^{\phi},\ \phi < 1\\
R_t &= s_tL_t = s_tL_0e^{nt},
\end{aligned}$$

where $h(K_t)$ is the function given in the main text in equation (32) and has the property that $\lim_{K\to\infty} h'(K) = 0$.

Combining some of these constraints, we see that the problem can be simplified further to

$$\max_{\{s_t\}} \int_0^\infty e^{-\theta t} u(c_t)\, dt \text{ subject to}$$
$$c_t = \frac{Y_t}{L_t} = \tilde{h}(N) N_t^{1/\beta}(1 - s_t)$$
$$\dot{N}_t = \alpha R_t^\lambda N_t^\phi, \phi < 1$$
$$R_t = s_t L_t = s_t L_0 e^{g_L t},$$

where $\tilde{h}(N) \equiv (\log 2)^{1/\beta} h(K(N))$ is asymptotically constant.

As N gets large, $\tilde{h}(N)$ converges to a constant and $\tilde{h}'(N) \to 0$, so this problem is a completely standard semiendogenous growth model. The optimal allocation features an allocation of research that converges to s^* such that

$$s^* = \frac{\psi}{1 + \psi}, \text{ where } \psi \equiv \frac{\lambda g_N/\beta}{\theta - (1 - \gamma) g_c + (1 - \phi) g_N},$$

where

$$g_y = g_c = \frac{g_N}{\beta} = \frac{1}{\beta}\frac{\lambda g_L}{1 - \phi}.$$

This is exactly the same as the solution in the main text in equation (33).

References

Acemoglu, Daron, and Pablo D. Azar. 2020. "Endogenous Production Networks." *Econometrica* 88 (1): 33–82.

Agrawal, Ajay, John McHale, and Alexander Oettl. 2019. "Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 149–74. Chicago: Univ. Chicago Press.

Akcigit, Ufuk, William R. Kerr, and Tom Nicholas. 2013. "The Mechanics of Endogenous Innovation and Growth: Evidence from Historical US Patents." Working paper, Univ. Chicago.

Akcigit, Ufuk, William R. Kerr, and Jeremy Pearce. 2022. "The Mechanics of Endogenous Innovation and Growth: Evidence from Historical US Patents." Working paper, Univ. Chicago.

Arnold, Barry C., Narayanaswamy Balakrishnan, and Haikady Navada Nagaraja. 2008. *A First Course in Order Statistics*. Philadelphia: Soc. Indus. Appl. Math.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293:1818–20.

Barlevy, Gadi. 2008. "Identification of Search Models Using Record Statistics." *Rev. Econ. Studies* 75 (1): 29–64.

Barton, D. E., and F. N. David. 1959. "Combinatorial Extreme Value Distributions." *Mathematika* 6 (1): 63–76.

Benassy, Jean-Pascal. 1996. "Taste for Variety and Optimum Production Patterns in Monopolistic Competition." *Econ. Letters* 52 (1): 41–47.

de Haan, Laurens, and Ana Ferreira. 2006. *Extreme Value Theory: An Introduction.* New York: Springer.

Eaton, Jonathan, and Samuel S. Kortum. 2002. “Technology, Geography, and Trade.” *Econometrica* 70 (5): 1741–79.

Economist. 2020. “How ASML Became Chipmaking’s Biggest Monopoly.” *Economist,* February 29.

Embrechts, Paul, Thomas Mikosch, and Claudia Klüppelberg. 1997. *Modelling Extremal Events: For Insurance and Finance.* Berlin: Springer.

Fleming, Lee, and Olav Sorenson. 2001. “Technology as a Complex Adaptive System: Evidence from Patent Data.” *Res. Policy* 30 (7): 1019–39.

Gabaix, Xavier. 1999. “Zipf’s Law for Cities: An Explanation.” *Q.J.E.* 114 (3): 739–67.

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

Glick, Ned. 1978. “Breaking Records and Breaking Boards.” *American Mathematical Monthly* 85 (1): 2–26.

Google Scholar. 2021. https://scholar.google.com.

Hall, Bronwyn H., and Rosemarie Ham Ziedonis. 2001. “The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979–1995.” *RAND J. Econ.* 32 (1): 101–28.

Hopenhayn, Hugo A. 1992. “Entry, Exit, and Firm Dynamics in Long-Run Equilibrium.” *Econometrica* 60 (5): 1127–50.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. 1995. “Extreme Value Distributions.” In *Continuous Univariate Distributions,* vol. 2, edited by Norman L. Johnson, Samuel Kotz, and N. Balakrishnan, chap. 22. Hoboken, NJ: Wiley.

Jones, Benjamin F. 2009. “The Burden of Knowledge and the ‘Death of the Renaissance Man’: Is Innovation Getting Harder?” *Rev. Econ. Studies* 76 (1): 283–317.

Jones, Charles I. 1995. “R&D-Based Models of Economic Growth.” *J.P.E.* 103 (4): 759–84.

———. 2023. “Replication Data for: ‘Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail.’” Harvard Dataverse, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/X0YWXF.

Jones, Charles I., and Jihee Kim. 2018. “A Schumpeterian Model of Top Income Inequality.” *J.P.E.* 126 (5): 1785–826.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. “Measuring Technological Innovation over the Long Run.” *A.E.R. Insights* 3 (3): 303–20.

Klenow, Peter J., and Jonathan L. Willis. 2016. “Real Rigidities and Nominal Price Changes.” *Economica* 83 (331): 443–72.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. “Technological Innovation, Resource Allocation, and Growth.” *Q.J.E.* 132 (2): 665–712.

Kortum, Samuel, and Josh Lerner. 1998. “Stronger Protection or Technological Revolution: What Is Behind the Recent Surge in Patenting?” *Carnegie-Rochester Conf. Ser. Public Policy* 48:247–304.

Kortum, Samuel S. 1997. “Research, Patenting, and Technological Change.” *Econometrica* 65 (6): 1389–419.

Lucas, Robert E., and Benjamin Moll. 2014. “Knowledge Growth and the Allocation of Time.” *J.P.E.* 122 (1): 1–51.

Luttmer, Erzo G. J. 2007. “Selection, Growth, and the Size Distribution of Firms.” *Q.J.E.* 122 (3): 1103–144.

———. 2010. "Models of Growth and Firm Heterogeneity." *Ann. Rev. Econ.* 2:547–76.

Martellini, Paolo, and Guido Menzio. 2020. "Declining Search Frictions, Unemployment, and Growth." *J.P.E.* 128 (12): 4387–437.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–725.

Pearce, Jeremy. 2022. "Idea Production and Team Structure." Working paper, Univ. Chicago.

Perla, Jesse, and Christopher Tonetti. 2014. "Equilibrium Imitation and Growth." *J.P.E.* 122 (1): 52–76.

Poincaré, Henri. (1910) 2000. "Mathematical Creation." *Resonance* 5:85–94. Original edition, *Monist* 20 (3): 321–35.

Resnick, Sidney I. 2008. *Extreme Values, Regular Variation, and Point Processes.* New York: Springer.

Romer, Paul M. 1990. "Endogenous Technological Change." *J.P.E.* 98 (5): S71–S102.

———. 1993. "Two Strategies for Economic Development: Using Ideas and Producing Ideas." *Proceedings of the World Bank Annual Conference on Development Economics 1992,* edited by Lawrence H. Summers and Shekhar Shah, 63–115. Washington, DC: World Bank.

Sadler, Matthew, and Natasha Regan. 2019. *Game Changer.* Alkmaar: New in Chess.

Scopus. 2021. https://www.scopus.com/.

US Patent and Trademark Office. 2020. "U.S. Patent Activity Calendar Years 1790 to the Present." https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm.

Web of Science. 2021. https://apps.webofknowledge.com.

Weitzman, Martin L. 1998. "Recombinant Growth." *Q.J.E.* 113 (2): 331–60.

Youn, Hyejin, Deborah Strumsky, Luis M. A. Bettencourt, and José Lobo. 2015. "Invention as a Combinatorial Process: Evidence from US Patents." *J. Royal Soc. Interface* 12 (106): 20150272.