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

Charles I. Jones*
Stanford GSB and NBER

July 5, 2020 — Version 0.5
Preliminary

Abstract

New ideas are often combinations of existing goods or ideas, a point emphasized by Romer (1993) and Weitzman (1998). A separate literature highlights the links between exponential growth and Pareto distributions: Gabaix (1999) shows how exponential growth generates Pareto distributions, while Kortum (1997) shows how Pareto distributions generate exponential growth. But this raises a "chicken and egg" problem: which came first, the exponential growth or the Pareto distribution? And regardless, what happened to the Romer and Weitzman insight that combinatorics should be an essential ingredient in understanding growth? This paper answers these questions by showing that combinatorial growth based on draws from standard thin-tailed distributions leads to exponential economic growth; no Pareto assumption is required.

*I am grateful to Daron Acemoglu, Pablo Azar, Sebastian Di Tella, Guido Imbens, Pete Klenow, Sam Kortum, and Chris Tonetti for helpful discussions about this project.

1. 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 5000 suppliers, including robotic arms and mirrors of unimaginable smoothness (The Economist, 2020). Romer (1993) 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 ideas that they are 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. Gabaix (1999), Luttmer (2007), and Jones and Kim (2018) emphasize that exponential growth, tweaked appropriately, can generate a Pareto distribution for city sizes, firm employment, or incomes. Conversely, Kortum (1997) shows that Pareto distributions are key to exponential growth: if productivity is the maximum over a number of draws from a distribution (you use only the best idea), then exponential growth in productivity requires that the number of draws grows exponentially and that the distribution being drawn from is Pareto, at least in the upper tail. Exponential growth and Pareto distributions, then, seem to be two sides of the same coin.

But this leads to a "chicken and egg" problem: which came first, the exponential growth or the Pareto distribution? And regardless, what happened to the Romer and Weitzman insight that combinatorics should be an essential ingredient in understanding growth?

This paper provides an answer to these questions by combining the insights of Kortum (1997) and Weitzman (1998). As in Kortum, we think of ideas as draws from some probability distribution. Building on Weitzman, we highlight a crucial role for combinatorics.

To see the insight, suppose 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, a total of 2^{N_t} possibilities. Finally, research consists of adding new recipes to the cookbook — i.e. evaluating them and learning their productivities. Specifically, we add the recipes associated with new ingredients to our cookbook according to $\dot{N}_t = \alpha L_t$, where L_t is the number of researchers. As t gets large, N_t grows exponentially with population growth. We call a setup with 2^N recipes with exponential growth in the number of ingredients *combinatorial growth*.

The key result in the paper is this: combinatorial expansion is so fast that drawing from a conventional thin-tailed distribution (e.g. a normal) 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 additional insights. For example, let K be the number of recipes in the cookbook and Z_K be the productivity of the best recipe. Let $\bar{F}(x)$ denote the probability that a recipe has a productivity *higher* than x — the complement of the cdf — so that it characterizes the search distribution. Then the key condition derived below that relates the growth in Z_K to the number of draws and the search distribution is this: Z_K grows asymptotically at the rate that makes $K\bar{F}(Z_K)$ stable. That is, given a rate of expansion of K, the maximum productivity marches down the upper tail of the distribution so as to render $K\bar{F}(Z_K)$ stationary. Kortum (1997) can be viewed in this context: exponential growth in Z_K is achieved by an exponentially growing number of draws K from a Pareto tail in $\bar{F}(\cdot)$. Similarly, combinatorial growth in K requires a tail that is an exponential of a power function. Even the Romer (1990) model can be viewed in this light: linear growth in K requires a log-Pareto tail for the search distribution. This same logic can essentially be applied to any setup: if you want exponential growth in Z_K from a particular search distribution $\bar{F}(\cdot)$, then you need the rate at which you take draws from the distribution to render $K\bar{F}(Z_K)$ stationary.

This perspective suggests a resolution of the "chicken and egg" problem mentioned above: exponential growth is the primitive and comes first. Economic growth does not require Pareto distributions. Then, through the logic suggested by Gabaix (1999) and Luttmer (2007), exponential growth can lead to Pareto distributions.

Section 2 below explains these basic insights in a simple setting, while Section 3 embeds the setup into a full growth model. Section 4 connects our results with the literature on extreme value theory to show how the results generalize to different distributions. In Section 5, we show that the model has an important empirical prediction: in the combinatorial case, the number of new ideas should be growing exponentially over time. This prediction provides a good description of the patent data in recent decades. We defer the literature review to the end of the paper in Section 6; several of the other important inspirations for this project — especially Acemoglu and Azar (2020) — are easier to discuss after we've laid out our framework.

2. Combining Weitzman and Kortum

As in Romer (1993), suppose there are a huge number of ingredients in the world that can be combined into ideas. This number is presumably finite, but Romer's point was that it is so large that the number of potential combinations is effectively infinite. Our cookbook, $\mathcal{C}$, is the set of all recipes we've evaluated as of some point in time. Let K denote the number of recipes in the cookbook.

Each recipe is an idea, and the idea 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 the value (productivity) associated with each recipe is an independent draw from some distribution. In particular, let z_c denote the value of recipe c and let $F(x)$ be the cumulative distribution function for each independent z_c. The only condition we make on $F(x)$ is that it is unbounded, continuous, and strictly increasing.

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. This is a simplified version of the Kortum (1997) setup. Let $Z_K \equiv \max z_c$

where $c = 1, ..., K$. Because we care about the best idea, it is convenient to define the tail probability (sometimes called the survival function):

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

From a growth theory standpoint, the question we are interested in 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 $F(x)$:

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

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. Intuitively, we need to "replace" the $\bar{F}(x)$ on the right side of (2) with something that depends on $1/K$ and then take the limit as K goes to infinity so that the exponential function appears.

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 (An alternative extreme value theorem)**.** *Let Z_K denote the maximum over K independent draws from an unbounded distribution with a strictly decreasing and continuous tail cdf $\bar{F}(x)$. Then*

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

Proof. Given that Z_K is the max over K i.i.d. draws, we have

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

Let $M_K \equiv K\bar{F}(Z_K)$ denote a new random variable. Then

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

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.[1] QED

Let's 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 $\bar{F}(Z_K)$ is getting smaller and smaller as we march down the tail of the distribution. On the other hand, multiplying by K raises the value away from zero. Theorem 1 says that under very weak conditions — basically that the underlying distribution we draw from is continuous — $K\bar{F}(Z_K)$ converges in distribution to a standard exponential distribution.

An alternative version of Theorem 1 is presented in Appendix A.1 that uses a Poisson assumption as in Kortum (1997) to derive the result at each point in time without needing to take the limit as t goes to infinity.

Intuitively, 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 of the distribution $\bar{F}(\cdot)$ decays tells us how the rates of increase of Z_K and K are related.

Let's now apply this logic to growth models, first as in Kortum (1997) and then in a new way involving combinatorics.

[1] This theorem must surely be known in the statistics literature already, but I do not have a reference. I discuss its relationship with the more standard extreme value theorem later in Section 4.

2.1 Kortum (1997)

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

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 therefore

$$\begin{aligned} \Pr\left[\, K\bar{F}(Z_K) \geq m \,\right] &= \Pr\left[\, KZ_K^{-\beta} \geq m \,\right] \\ &= \Pr\left[\, K^{-1/\beta} Z_K \leq m^{-1/\beta} \,\right] \end{aligned} \tag{5}$$

and which in turn equals e^{-m} in the limit by Theorem 1. Now call the cutoff point in (5) that we are considering $x \equiv m^{-1/\beta}$ so that $m = x^{-\beta}$. This definition with Theorem 1 and equation (5) gives

$$\boxed{\lim_{K\to\infty} \Pr\left[\, K^{-1/\beta} Z_K \leq x \,\right] = e^{-x^{-\beta}}.} \tag{6}$$

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 as a Fréchet distribution, also known as the Type II extreme value distribution.

Letting ε be a draw from this Fréchet distribution, equation (6) implies that for K large,

$$Z_K \approx K^{1/\beta}\varepsilon$$

If the number of draws K grows exponentially at rate g_L (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

$$g_Z = \frac{g_L}{\beta}. \tag{7}$$

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

2.2 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's 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's assume that our cookbook consists of all recipes that come from combining N ingredients. Each ingredient can either be included or excluded from a recipe, so there are a total of $K = 2^N$ recipes that can be made from N ingredients. At a given point in time, the economy picks from $K = 2^N$ different combinations and chooses the recipe that is best.

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

$$\begin{aligned}\Pr\left[\, K\bar{F}(Z_K) \geq m \,\right] &= \Pr\left[\, Ke^{-\theta Z_K} \geq m \,\right] \\ &= \Pr\left[\, \log K - \theta Z_K \geq \log m \,\right] \\ &= \Pr\left[\, Z_K - \frac{1}{\theta}\log K \leq -\frac{1}{\theta}\log m \,\right] \end{aligned} \tag{8}$$

and which in turn equals e^{-m} in the limit. Now redefine the cutoff point to be $x \equiv -(1/\theta)\log m$ so that $m = e^{-\theta x}$ and combine this change of variables with Theorem 1 and equation (8) to get

$$\boxed{\lim_{K\to\infty} \Pr\left[Z_K - \frac{1}{\theta}\log K \le x\right] = e^{-e^{-\theta x}}.} \tag{9}$$

That is, to get a stable distribution for the max over K draws from an exponential distribution, we subtract $(1/\theta)\log K$ from the max Z_K. This appropriately-scaled max then is distributed asymptotically as a Gumbel distribution, also known as the Type I extreme value distribution.

Letting ε be a draw from this Gumbel distribution, equation (9) implies that for K large,

$$Z_K \approx \frac{1}{\theta}\log K + \varepsilon \tag{10}$$

If the number of draws K were to grow exponentially at rate g_L, say because of population growth in the number of researchers, 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 in this paper is that if the number of draws is combinatorial instead, it is possible to restore exponential growth. In particular if $K = 2^N$ and N grows exponentially at rate g_L, then

$$Z_K \approx \frac{1}{\theta}\log 2^N + \varepsilon = \frac{1}{\theta}N\log 2 + \varepsilon \tag{11}$$

and the asymptotic growth rate of productivity in this economy will average to

$$g_Z = g_L. \tag{12}$$

Productivity growth is asymptotically equal to the growth rate of the number of ingredients whose recipes have been evaluated, which equals the growth rate of researchers (if, say, $\dot{N}_t = \alpha L_t$).

The key new growth result is then 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 a thin-tailed distribution to generate exponential growth in productivity.

2.3 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 4, we will see even more generalizations.

Equation (10) implies that

$$\frac{Z_K}{\log K} \xrightarrow{p} \text{Constant}\,. \tag{13}$$

That is, the ratio of the max from K draws of an exponential to $\log K$ converges in probability to a constant. (This is shown more formally in Section 4.2 below.)

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:

$$\begin{aligned} \frac{\max y}{\log K} &\xrightarrow{p} \text{Constant} \\ \Rightarrow \quad \frac{\max x^{\beta}}{\log K} &\xrightarrow{p} \text{Constant} \\ \Rightarrow \quad \frac{\max x}{(\log K)^{1/\beta}} &\xrightarrow{p} \text{Constant} \end{aligned} \tag{14}$$

where here and later we will follow the convention that "Constant" denotes an unimportant constant that may change across equations. 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_L, the growth rate of the max is asymptotically given by

$$g_Z^{\textit{weibull}} = \frac{g_L}{\beta} \tag{15}$$

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 researchers 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).

3. Growth Model

The economic environment for the full growth model is shown in Table 1. Aggregate output is a 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 (typically different) recipe from the cookbook. A recipe uses M_{it} 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-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 by its productivity index, z_{ic}. As in the statistical model above, the z_{ic}'s for each recipe are i.i.d. draws from a common distribution, which we assume for now is Weibull; in the next section, we will explain how this generalizes. At any given point in time, the cookbook contains K_t recipes that have been evaluated, and the firm producing each variety chooses the recipe with the highest productivity, Z_{Ki}.

The evolution of recipes in the cookbook follows a *combinatorial growth process*, as defined earlier. We generalize it slightly to incorporate intertemporal spillovers: with R_t 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 ($\phi < 0$). Because of combinatorics, the number of recipes in the cookbook at each point in time is $K_t = 2^{N_t}$, where N_t is the number of ingredients that have been evaluated as of date t.

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, L_t. And there is exponential population growth at constant rate g_L.

3.1 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:

Table 1: The Economic Environment

Aggregate output	$Y_t = \left(\int_0^1 Y_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}$ with $\sigma > 1$	(16)
Variety i output	$Y_{it} = Z_{Kit} \left(M_{it}^{-\frac{1}{\rho}} \sum_{j=1}^{M_{it}} x_{ijt}^{\frac{\rho-1}{\rho}} di \right)^{\frac{\rho}{\rho-1}}$ with $\rho > 1$	(17)
Production of ingredients	$x_{ijt} = L_{ijt}$	(18)
Best recipe	$Z_{Kit} = \max_c z_{ic}, \;\; c = 1, ..., K_t$	(19)
Weibull distribution of z_{ic}	$z_{ic} \sim F(x) = 1 - e^{-x^{\beta}}$	(20)
Number of ingredients evaluated	$\dot{N}_t = \alpha R_t^{\lambda} N_t^{\phi}, \quad \phi < 1$	(21)
Cookbook	$K_t = 2^{N_t}$	(22)
Resource constraint: workers	$L_{it} = \sum_{j=1}^{M_i} L_{ijt}$ and $\int_0^1 L_{it} di = L_{yt}$	(23)
Resource constraint: R&D	$R_t + L_{yt} = L_t$	(24)
Population growth (exogenous)	$L_t = L_0 e^{g_L t}$	(25)

$R_t = \bar{s} L_t$.

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

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

Subsituting 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

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

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

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

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

Turning to the research side of the model,

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

and therefore as $t \to \infty$ we have

$$g_N = \frac{\lambda g_L}{1 - \phi}. \tag{30}$$

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.

From equation (14),

$$\frac{Z_{Kit}}{(\log K_t)^{1/\beta}} \xrightarrow{p} \text{Constant} \Rightarrow \frac{Z_{Kt}}{(\log K_t)^{1/\beta}} \xrightarrow{p} \text{Constant} \tag{31}$$

and therefore[2]

$$\boxed{g_y = g_{Z_K} = \frac{g_N}{\beta} = \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}.} \tag{32}$$

As was suggested by the basic statistical model, we have a setting where output per person, $y \equiv Y/L$, grows exponentially. Superior 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 in the cookbook, driven by population growth in the number of researchers, offsets the thinness of the tail in the search distribution 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" via $\phi > 0$.

4. Generalizing to other distributions

In the previous sections, 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.

4.1 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 $\frac{Z_K - b_K}{a_K}$ 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, all we assumed essentially is that the distribution function is continuous and invertible.

[2]The last step in the preceding equation is shown in more detail in Appendix A.3.

An intuitive way to see why additional assumptions are not required is this: Because Z_K is a random variable, $\bar{F}(Z_K)$ is also a random variable. In particular, $\bar{F}(Z_K)$ is distributed Uniform on $(0,1)$, and this is true regardless of the particular distribution. Since Z_K is the max from $F(x)$ and since $\bar{F}(x)$ is a decreasing function, $\bar{F}(Z_K)$ is the minimum over K draws from a $U(0,1)$. In this interpretation, equation (3) of Theorem 1 just says that K times the minimum of K draws from a $U(0,1)$ is asymptotically distributed as an exponential. This narrow result is a well-known in statistics. But it has broad implications for extreme value theory, as we show in what follows.

4.2 Scaling and Growth for Other Distributions

Now let's see how the results generalize to other distributions. First, rewrite equation (3) as

$$K\bar{F}(Z_K) = \varepsilon + o_p(1) \tag{33}$$

where ε is a random variable from an exponential distribution with parameter equal to one. One way to proceed is to plug in different distribution functions and derive the scaling. In particular, we will derive expressions for b_K such that

$$\frac{Z_K}{b_K} \xrightarrow{p} \text{Constant}\,.$$

To start, return to the exponential, $\bar{F}(x) = e^{-\theta x}$. In this case, (33) implies

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

and therefore

$$\frac{Z_K}{\log K} \xrightarrow{p} \text{Constant}\,. \tag{34}$$

That is, Z_K grows with $\log K$ in the exponential case. This is just a more formal way of expressing the result we derived earlier in equation (10).

This same line of reasoning can be used to derive the scaling for other distributions, and we will use it one more time a bit later in providing microfoundations for Romer (1990). However, there is a convenient "change of variables" shortcut that works for many distributions. We already introduced this shortcut above in Section 2.3 to derive the scaling result for the Weibull distribution, namely that $Z_K/(\log K)^{1/\beta} \xrightarrow{p} \text{Constant}$.

The change-of-variables method does not directly work when the search distribution is a normal distribution. For that case, the standard approach of extreme value theory gives the scaling. This is discussed further in Appendix A.2, but the result is that the maximum scales with $b_K = (\log K)^{1/2} = \sqrt{\log K}$. The parallel to the Weibull distribution is instructive: the tail of a normal falls with e^{-x^2} and the exponent in b_K is 1/2; the tail of a Weibull falls with e^{-x^β} and the exponent in b_K is $1/\beta$.

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

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

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.[3]

[3]To see another interesting application of this fact, suppose $\log x$ is drawn from the exponential distribution. But this means that x is 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.

However, one can calculate what growth rate of K 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. It is well known (and easy to show using Theorem 2 in Appendix A.2) that the Gompertz distribution is in the Gumbel domain of attraction. Then the change-of-variables method works here: assume y is exponentially distributed, and let $y = e^{\beta x} - 1$ so that x has a Gompertz distribution. Then

$$\begin{aligned}
\frac{\max y}{\log K} &\xrightarrow{p} \text{Constant} \\
\Rightarrow \quad \frac{\max e^{\beta x} - 1}{\log K} &\xrightarrow{p} \text{Constant} \\
\Rightarrow \quad \frac{\max e^{\beta x}}{\log K} &\xrightarrow{p} \text{Constant} \\
\Rightarrow \quad \frac{\max x}{\frac{1}{\beta}\log(\log K)} &\xrightarrow{p} \text{Constant}
\end{aligned}$$

In this case, the max grows with $\log(\log K)$. Exponential growth in the max requires $\log(\log K)$ to grow exponentially. Even combinatoric 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, in the sense that $\bar{F}(x)/\bar{G}(x) \to$ Constant:

$$\lim_{x\to\infty} \frac{e^{-x}}{1 - e^{-e^{-x}}} = 1.$$

That is, for x large, $e^{-e^{-x}} \approx 1 - e^{-x}$, so the Gumbel has an exponential upper tail. For this reason, it also features $b_K = \log K$, just like the exponential.

Microfoundations for Romer (1990). There is a final special case worth considering. One of the key 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 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 (33):

$$\begin{aligned} K\bar{F}(Z_K) &= \varepsilon + o_p(1) \\ \Rightarrow \quad \frac{K}{(\log Z_K)^\alpha} &= \varepsilon + o_p(1) \\ \Rightarrow \quad \frac{\log Z_K}{K^{1/\alpha}} &= \left(\frac{1}{\varepsilon + o_p(1)}\right)^{1/\alpha} \end{aligned} \tag{35}$$

Next, if ε is distributed as exponential with parameter one, then $\varepsilon^{-1/\alpha}$ is a Fréchet random variable with parameter α.[4] Using this fact in equation (35) gives

$$\frac{\log Z_K}{K^{1/\alpha}} \overset{a}{\sim} \text{Fréchet}(\alpha) \tag{36}$$

[4] Since ε has an exponential distribution with parameter equal to one,

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

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

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

and therefore

$$Z_K = \exp\left(K^{1/\alpha}\right)\exp(\tilde{\varepsilon} + o_p(1)) \tag{37}$$

where $\tilde{\varepsilon}$ is a Fréchet random variable with parameter α. We now have the scaling: the maximum over draws from a log-Pareto distribution grows asymptotically with $\exp(K^{1/\alpha})$.

To see the microfoundations for Romer (1990), suppose $\dot{K}_t = \beta L$ where L is a constant population. Then $K(t) = K_0 + gt$ grows linearly where $g \equiv \beta L$ and — if $\alpha = 1$ — the max will grow asymptotically as $\exp(gt)$. In other words, if our productivity draws are log-Pareto distributed with the Pareto parameter equal to one (so that even the mean of the Pareto distribution does not exist), 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 (35), 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.

Summary. These results are collected together in Table 2. In particular, they show how the number of draws from the search distribution, K_t, must behave in order to generate exponential growth in Z_K for different distributions. That is, they show how to stabilize $K\bar{F}(Z_K)$. There is a tradeoff between the shape of the tail of the search distribution and the rate at which we march down that tail.

In order for combinatorial growth to deliver exponential growth in the maximum, we need the max to grow with $(\log K)^{1/\beta}$, i.e. as a power function of the log of the number of draws. Distributions in which the tail is asymptotically equivalent to an exponential of a power function — the Weibull being a canonical example — deliver this result. Examples include the exponential, the Gumbel, and the normal distributions, but Embrechts, Mikosch and Klüppelberg (1997) provide other examples as well, including generalizations of the Weibull distribution (e.g. where $\bar{F}(x) = x^{\alpha}e^{-x^{\beta}}$), the gamma distribution, and the Benktander Type I and Type II distributions. If instead the $\log x$ is drawn from one of these distributions, the tail will be too thick and combinatorial growth will explode. Alternatively, if the tail falls off as the exponential of

Table 2: Scaling of Z_K for Various Distributions

Distribution	cdf	b_K	$b_K(N)$ for for $K = 2^N$	Growth rate for $K = 2^N$
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: The maximum over K i.i.d. draws from a distribution in the Gumbel domain of attraction scales asymptotically with the normalizing sequence b_K (here we are ignoring multiplicative constants — for example, $1/\theta$ in the exponential case). If b_K grows exponentially, then Z_K will as well. The last two columns focus on the combinatorial case. The penultimate column translates this into scaling with N for $K = 2^N$ (ignoring some multiplicative constants). The final column shows the asymptotic growth rate of Z_K if $N(t)$ grows exponentially at rate g_N.

an exponential function (as in the Gompertz case), then the tail will be too thin for combinatorial draws to deliver exponential growth.

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.

The productivity of each recipe can be drawn from a normal, Weibull, exponential, gamma, logistic, or Gumbel distribution — or indeed any distribution that has a thin tail in the sense that it decays exponentially as a polynomial function of x. In all of these cases, the maximum over K draws will rise with $\log K = \log 2^N$. Therefore, if the number of ingredients being evaluated rises exponentially, all of these cases will lead to exponential growth. *Combinatorial expansion with draws from many common thin-tailed distribution generates exponential growth.*

5. Evidence

One of the facts that Kortum (1997) sought to address 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 U.S. inventors in 1915, 1950, and 1985 was roughly the same, around 40,000. In his setup, each new idea is endogenously a proportional improvement on the previous state-of-the-art, so that a constant flow of new ideas can generate exponential growth.

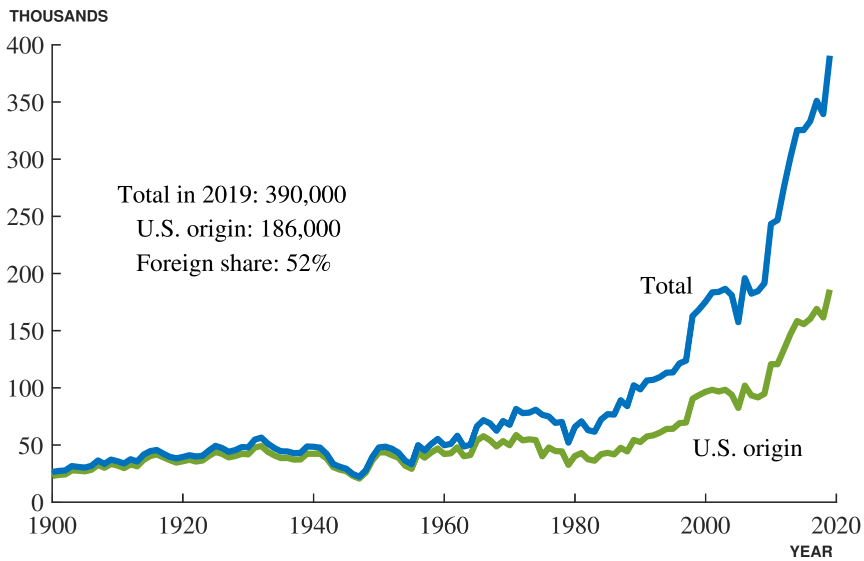
However, even at the time he was writing, this fact was already changing. Figure 1 shows the time series for patents granted by the U.S. Patent Office, both in total (i.e. including foreign inventors) and to U.S. inventors only. Far from being constant, the patent series viewed from the perspective of 2020 looks much more like a series that itself exhibits exponential growth. Put differently, the rise in patents in the United States would, in Kortum (1997), imply a substantial increase in the rate of economic growth, something we don't see empirically.

One resolution of this discrepancy 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; 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.

Alternatively, perhaps the series for new ideas is in fact growing exponentially over time, as suggested by Figure 1. The interesting observation I want to put forward in the remainder of this section is that this is precisely what the combinatorial growth model predicts.

To see this point, we first have to define what we mean by a patent or a new idea in

Figure 1: Patents Granted by the U.S. Patent and Trademark Office



Source: U.S. Patent and Trademark Office (2020).

the model. We follow Kortum (1997) in defining patents or new ideas to be ideas that are improvements over the state-of-the-art. If there are K_t recipes in the cookbook, how many of them exceeded the "state-of-the-art" when they were discovered?

The theory of record breaking suggests the following simple insight. If the draws are independent, then the probability that any one of the K_t recipes is the best is just $1/K_t$. 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 simply $\dot{K}_t/K_t$. 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.

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 $\phi = 0$).

$$\begin{aligned} K_t &= 2^{N_t} \\ \Rightarrow \frac{\dot{K}_t}{K_t} &= \log 2 \cdot \dot{N}_t \\ &= \log 2 \cdot \alpha \bar{s} L_t \\ &= \log 2 \cdot \alpha \bar{s} L_0 e^{g_L t} \end{aligned} \tag{38}$$

That is, the number of 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.

The intuition for this result is straightforward: because of the thin tail of the probability distribution, the typical new idea is only slightly better than the previous state-of-the-art. Exponential growth in productivity requires us to march down the tail very quickly — combinatorially — and this delivers exponential growth in the number of "patents" in the model. The growth that we see empirically in the actual patent series, then, is potentially evidence for the combinatorial growth process itself.

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 two responses to this question. The first is the empirical evidence provided above: the combinatorial process leads to exponential growth in patenting, 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 a combinatorial number of moves 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 (38) above, so perhaps this is not as implausible as it at first appears.

6. Discussion and Further Connections to the Literature

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

Acemoglu and Azar (2020). 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 recipes be a draw from a probability distribution. Their setup inspired the approach taken in this paper.

Where the two papers go in different directions is 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.[5] Their Corollary 2 suggests that broader results are possible with different growth rates for the number of new goods, and this paper can be interpreted as exploring those broader results.

New ideas as new ingredients? To what extent are new ideas themselves new ingredients that can be used in future recipes? We made a conscious decision early on in this paper to follow Weitzman (1998)'s lead in emphasizing that there are large numbers of potential ideas and growth is limited by our ability to evaluate the merits of those ideas. In this sense, the *evaluation* equation $\dot{N}_t = \alpha R_t^\lambda N_t^\phi$ and the size of the cookbook 2^{N_t} do not change just because new ideas are themselves potential new ingredients that can be tried. As in Weitzman, there are so many potential ideas that processing and evaluation are the key limits. An alternative approach one could take, however, is to say

[5] 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.

the number of ingredients is initially small and that the new ideas are themselves new ingredients. This approach can lead to faster-than-combinatorial expansion, more like the "towers" of $2^{2^{2^{\cdots}}}$. Ultimately, this is just another reason why our ability to evaluate ideas is the decisive constraint.

A somewhat related concern is that of 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. What is clear from the paper, however, is that if you want exponential growth from draws from a distribution with an exponential tail, you will need the *effective* number of draws, i.e. taking the correlation into account, to exhibit combinatorial growth.

Models of technology diffusion. A potentially 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 similar 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. Because of revolutions in communication technologies, it is arguable that the diffusion of ideas occurs much faster today than in the past. Perhaps combinatorial diffusion plus thinned-tailed distributions can be applied in this setting as well.

Pareto and the chicken-and-egg problem. Finally, as discussed in the Introduction, one of the motivations for this project was the "chicken-and-egg" aspect of exponential growth and Pareto distributions. We do seem to see Pareto distributions empirically in many places, including the size of cities, the size of firms, and the income and wealth distributions. The resolution suggested here is that exponential growth comes first. Then the mechanism of Gabaix (1999) and Luttmer (2007) that exponential growth can be used to generate Pareto distributions is a candidate explanation for the Pareto distributions that we see in the data. It would be interesting to micro-found this story using the combinatorial process presented here.

Conclusion. In the end, the paper can be read in two ways. First, there is the "Weitzman meets Kortum / combinatorial growth" interpretation: if we have the number of draws growing 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. In considering 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 take any strictly monotonic continuous distribution $\bar{F}(x)$ and reverse engineer the time path for K that is required to generate exponential growth in Z_K.

A. Appendix

A.1 Theorem 1 in the Poisson Case

Here we state a version of Theorem 1 that uses a Poisson assumption to get the extreme value result for all t rather than as an asymptotic result. This follows the approach taken in Kortum (1997). I am grateful to Sam Kortum for suggesting it and providing the derivation.

Corollary 1 (Poisson version of Theorem 1)**.** *Let Z_K denote the maximum over K independent draws from an unbounded distribution with a strictly decreasing and continuous tail cdf $\bar{F}(x)$ and suppose K is distributed as Poisson with parameter T. Then*

$$\Pr\left[\, T\bar{F}(Z_K) \geq y \,\right] = e^{-y}. \tag{39}$$

Proof. Given that Z_K is the max over K i.i.d. draws, we have

$$\Pr\left[\, Z_K \leq x \,\right] = (1 - \bar{F}(x))^K. \tag{40}$$

Let $Y_K \equiv T\bar{F}(Z_K)$ denote a new random variable, conditional on K. Then

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

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 (40).

Now, we use the Poisson assumption to get the unconditional distribution of Y:

$$\begin{aligned}
\Pr\left[\,Y \geq y\,\right] &= \sum_{K=0}^{\infty} \Pr\left[\,Y_K \geq y\,\right] \cdot \Pr\left[\,K|\,T\,\right] \\
&= \sum_{K=0}^{\infty} \left(1 - \frac{y}{T}\right)^K \cdot \frac{e^{-T}T^K}{K!} \\
&= e^{-y} \sum_{K=0}^{\infty} \frac{e^{-T(1-y/T)}(T(1-y/T))^K}{K!} \\
&= e^{-y}
\end{aligned}$$

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 $T(1-y/T)$, i.e., the value of the CDF at infinity which is equal to one. QED

As in Kortum (1997), this approach could be used to derive growth results that apply at each point in time rather than asymptotically.

A.2 Extreme Value Theory

Like the Central Limit Theorem, the Extreme Value Theorem is quite general. In particular, it says that if the asymptotic distribution of the normalized maximum over K i.i.d. random variables exists, then it takes one of three forms: Fréchet, Gumbel, or a bounded distribution. The bounded case occurs when the draws themselves are from a distribution that is bounded from above, which is not especially interesting from a growth standpoint, so we will ignore that case. The other two have already been suggested by the examples in the main text. Here, we note how those examples generalize. These points are explored in great detail by Galambos (1978), Johnson, Kotz and Balakrishnan (1995), Embrechts, Mikosch and Klüppelberg (1997), and de Haan and Ferreira (2006).

The tail characteristics of the $F(x)$ distribution determine whether the normalized maximum has a Fréchet or a Gumbel distribution. If tail probability $\bar{F}(x)$ declines as a power function (polynomial function), then the normalized max converges to a Fréchet distribution. Examples of distributions that satisfy this condition are the Pareto, the Cauchy, the Student t, and the Fréchet distribution itself.[6]

Alternatively, if $\bar{F}(x)$ declines as an exponential function, then the normalized max has a Gumbel distribution. Many familiar unbounded distributions fall into this category: the normal, lognormal, exponential, Weibull, Gompertz, logistic, and gamma distributions, as well as the Gumbel distribution itself. These distributions feature a wide range in terms of the thickness of the upper tail.

The extreme value theorem for distributions in the domain of attraction of the Gumbel distribution can be stated as follows, using definitions we've already provided.

Theorem 2. *Consider the unbounded distribution $F(x)$, and let Z_K be the maximum over K i.i.d. draws from the distribution. Define $h(x) = (1 - F(x))/F'(x) = \bar{F}(x)/F'(x)$ to be the inverse hazard function. If $lim_{x\to\infty} h'(x) = 0$, then there exist normalizing sequences $a_K > 0$ and b_K such that*

$$\lim_{K\to\infty} \Pr\left[\frac{Z_K - b_K}{a_K} \leq x\right] = e^{-e^{-x}}. \tag{41}$$

Furthermore, let $U(t)$ be defined as the inverse function of $1/(1 - F(x))$. Then the normalizing sequences a_K and b_K can be chosen as $b_K = U(K)$ and $a_K = KU'(K) = 1/(KF'(b_K))$.

Proof. This is just a restatement of (a simplified version of) Theorem 1.1.8 in de Haan and Ferreira (2006).

Some remarks about this theorem. First, the function $h(x)$ is just a scaled version

[6] Example 1.3.3 of Galambos (1978) considers $F(x) = 1 - 1/\log(x)$. Notice that this tail falls off more slowly than a power function. It has a thicker tail even than a Pareto distribution with parameter value 1, for which the mean fails to exist. The distribution of the normalized maximum fails to converge in this case. Galambos calculates that the maximum over just four draws from this distribution has a greater than 20 percent probability of being larger than 60 million!

of the probability that the draws are above x. If this tail probability falls to zero sufficiently quickly, then the normalized maximum asymptotically has a standard Gumbel distribution. Written differently,

$$\frac{Z_K - b_K}{a_K} \overset{a}{\sim} \text{Gumbel} \tag{42}$$

Letting ε be a random variable from a standard Gumbel distribution, equation (42) is equivalent to

$$Z_K = b_K + a_K\varepsilon + o_p(a_K). \tag{43}$$

Dividing both sides by b_K,

$$\frac{Z_K}{b_K} = 1 + \frac{a_K}{b_K} \cdot \varepsilon + \frac{o_p(a_K)}{b_K}.$$

Finally, it can be shown that $\lim_{K\to\infty} a_K/b_K = 0$ according to Embrechts, Mikosch and Klüppelberg (1997).[7] Therefore, we have the important result that

$$\boxed{\frac{Z_K}{b_K} \xrightarrow{p} 1.} \tag{44}$$

That is, the ratio of the max to b_K converges in probability to the value one. Asymptotically, in other words, the max grows just like the normalizing sequence $b_K = U(K)$. To understand the growth of the max, then, we just need to understand $b_K = U(K)$.

Table 3.4.4 of Embrechts, Mikosch and Klüppelberg (1997) reports the b_K (which is d_n in their notation) for many distributions, including the normal distribution discussed in the main text.

[7]See p. 149 and p. 141, noting that their notation is c_n/d_n; it is easy to verify for example distributions in their Table 3.4.4.

A.3 Deriving Equation (31) in Section 3

For a continuum of sectors in equation (29) and using (43) above, we have

$$\begin{aligned} Z_K &= \left(\int_0^1 Z_{Kit}^{\sigma-1} di \right)^{\frac{1}{\sigma-1}} \\ &= \left(\int_0^1 (b_K + a_K \varepsilon_i + o_p(a_K))^{\sigma-1} di \right)^{\frac{1}{\sigma-1}} \end{aligned}$$

Dividing by b_K:

$$\frac{Z_K}{b_K} = \int_0^1 (1 + \frac{a_K}{b_K}\varepsilon_i + \frac{o_p(a_K)}{b_K})^{\sigma-1} di \Big)^{\frac{1}{\sigma-1}} \tag{45}$$

$$\xrightarrow{p} 1. \tag{46}$$

References

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

de Haan, Laurens and Ana Ferreira, *Extreme Value Theory: An Introduction (Springer Series in Operations Research and Financial Engineering)*, Springer, 2006.

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

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

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

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan, "Chapter 22. Extreme Value Distributions," in "Continuous Univariate Distributions, Volume 2," Wiley Interscience, 1995.

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

— and Jihee Kim, "A Schumpeterian Model of Top Income Inequality," *Journal of Political Economy*, October 2018, *126* (5), 1785–1826.

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

Lucas, Robert E. and Benjamin Moll, “Knowledge Growth and the Allocation of Time,” *Journal of Political Economy*, February 2014, *122* (1), 1–51.

Luttmer, Erzo G.J., “Selection, Growth, and the Size Distribution of Firms,” *Quarterly Journal of Economics*, 08 2007, *122* (3), 1103–1144.

Perla, Jesse and Christopher Tonetti, “Equilibrium Imitation and Growth,” *Journal of Political Economy*, February 2014, *122* (1), 52–76.

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

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

— , “Two Strategies for Economic Development: Using Ideas and Producing Ideas,” *Proceedings of the World Bank Annual Conference on Development Economics, 1992*, 1993, pp. 63–115.

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

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

U.S. Patent and Trademark Office, “U.S. Patent Activity Calendar Years 1790 to the Present,” 2020. https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm, accessed June 10, 2020.

Weitzman, Martin L., “Recombinant Growth,” *Quarterly Journal of Economics*, May 1998, *113*, 331–360.