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

Charles I. Jones*
Stanford GSB and NBER

July 1, 2021 — Version 2.0

**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 important? This paper answers these questions by demonstrating that combinatorial growth in the number of draws from standard thin-tailed distributions leads to exponential economic growth; no Pareto assumption is required. More generally, it provides a theorem linking the behavior of the max extreme value to the number of draws and the shape of the tail for continuous probability distributions.

*I am grateful to Daron Acemoglu, Pablo Azar, Sebastian Di Tella, Brian Cevallos Fujiy, Guido Imbens, Pete Klenow, Sam Kortum, Erzo Luttmer, Ben Moll, Chris Tonetti, Tom Winberry, and seminar participants at Cornell, the Harvard Growth Lab, the LSE, the Minneapolis Fed, Peking University, Stanford, UAB Barcelona, UC Riverside, and Wharton for helpful discussions and comments and to Jack Hirsh for excellent research assistance.

## 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 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. 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 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. 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 central to understanding growth?

This paper answers 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: if each ingredient can either be 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 — i.e. 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*.

One key result in the paper is this: combinatorial expansion 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 $K$ outcomes. Let $\bar{F}(x)$ denote the probability that a draw has a productivity *higher* than $x$ — the complement of the 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: linear growth in $K$ requires a log-Pareto tail for the search distribution if the max is to exhibit exponential growth.

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 a Pareto assumption but can be obtained from combinatorial expansion with standard thin-tailed distributions. Then, through the logic suggested by Gabaix (1999) and Luttmer (2007), exponential growth can generate the Pareto distributions we observe.[1]

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 shows that the annual flows of academic publications and patents, both in the aggregate and by technology class, have risen sharply over time, supporting the combinatorial model.

The remainder of the paper is organized as follows. Section 2 below explains these basic insights in a simple setting, while Section 3 embeds the setup in a full growth model. Section 4 connects our results with the literature on extreme value theory and shows how the results generalize to different distributions. Section 5 presents the evidence on patents and publications, providing empirical support for the model. We defer a further review of the literature 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

Suppose 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've evaluated as of some point in time. Let $K$ denote the number

[1] Other resolutions to the "chicken and egg" problem are possible, of course: any theory of exponential growth that doesn't rely on Pareto distributions can qualify, such as Aghion and Howitt (1992) or Luttmer (2015). What is new here is explaining how to do this in the class of models that involves marching down the tail of some probability distribution.

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 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 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, as in Kortum (1997). Let $Z_K \equiv \max z_c$ where $c \in \{1, ..., K\}$. Because we care about the best idea, it is convenient to define the tail probability (sometimes called the survival function):

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

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,

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

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, 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** (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 $m \geq 0$*

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

*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 for $0 \leq m < K$

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

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. 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 $\bar{F}(Z_K)$ — 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 $\varepsilon$ is an exponential random variable with a mean equal to one. This version helps make apparent the sense in which the increases in $K$ and $Z_K$ offset in a way that is mediated by the tail $\bar{F}(\cdot)$ of the underlying distribution.

Second, $K\bar{F}(Z_K)$ is a measure of "luck relative to trend." Asymptotically, it has a mean equal to one. Values bigger than one suggest that $\bar{F}(Z_K)$ is high relative to $K$, so that $Z_K$ is unexpectedly low given the number of draws. Values less than one similarly suggest that the max $Z_K$ is surprisingly high. On average, though, the luck cancels out.

Third, 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 3 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, Chapter 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 4.

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

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

(1997) and then in a new way involving combinatorics.

## 2.1 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.

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 $\beta$ 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

$$\boxed{\frac{Z_K}{K^{1/\beta}} = (\varepsilon + o_p(1))^{-1/\beta}.} \tag{6}$$

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 just like $\tilde{\varepsilon} \equiv \varepsilon^{-1/\beta}$, which has a Fréchet distribution. For $K$ large,

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

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

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

It equals the growth rate of the number of draws deflated by $\beta$, 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. (Recall that $2^N = \sum_{k=0}^{N} \binom{N}{k}$, the total number of combinations.) At a given point in time, the economy picks from $K = 2^N$ different combinations and chooses the recipe that is best. We say $K$ 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

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

and therefore

$$\boxed{\frac{Z_K}{\log K} \xrightarrow{p} \text{Constant}} \tag{8}$$

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 $g_K$, 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, exponential growth is restored. In particular if $K = 2^N$ and $N$ grows exponentially at rate $g_N$, then

$$\frac{Z_K}{\log K} = \frac{Z_K}{N \log 2} \xrightarrow{p} \text{Constant} \tag{9}$$

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

$$g_Z = g_{\log K} = g_N. \tag{10}$$

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.

## 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 derive a necessary and sufficient condition for combinatorial draws to generate exponential growth, precisely characterizing the generality.

Equation (8) 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:

$$\frac{\max y}{\log K} \xrightarrow{p} \text{Constant}$$
$$\Rightarrow \frac{\max x^\beta}{\log K} \xrightarrow{p} \text{Constant}$$
$$\Rightarrow \frac{\max x}{(\log K)^{1/\beta}} \xrightarrow{p} \text{Constant} \tag{11}$$

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

$$g_Z^{weibull} = \frac{g_N}{\beta} \tag{12}$$

Intuitively, a higher value of $\beta$ 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 $\beta$, 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

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.

## 3.1 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 $P$ is distributed as Poisson with parameter $K$.*

*Then for* $0 \le m < K$ *and when* $P > 0$ *(so there are observations over which to take the max)*

$$\Pr\left[\, K\bar{F}(Z_K) \ge m \,\right] = \frac{e^{-m} - e^{-K}}{1 - e^{-K}}. \tag{13}$$

*Proof.* See Appendix A.2. QED

In the corollary, notice that the $e^{-K}$ term appears because $\Pr\left[\, P = 0 \,\right] = e^{-K}$ and $\Pr\left[\, P > 0 \,\right] = 1 - e^{-K}$ — the max only exists 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 and derivation of several of the later results.

### 3.2 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 simple continuous-time growth framework.

Aggregate output is a CES combination of a unit measure of varieties, as in equation (14). The production of each variety is given by (15). Each variety is produced using a different recipe from the cookbook. A recipe uses $M_{it} \le N_t$ ingredients that combine in a CES fashion, and one unit of each ingredient can be produced with one worker, as in equation (16). The $M_{it}^{-1/\rho}$ term in (15) is a Benassy (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, $z_c$.

We assume the productivity of recipes in the cookbook is revealed as a Poisson process. In particular, the flow of recipes that are learned between date $s$ and date $t$ is Poisson with parameter $K_t - K_s$. Because of the additivity of the Poisson process, the total number of recipes in the cookbook as of date $t$ is Poisson with parameter $K_t = 2^{N_t}$. A new recipe applies to one of the unit measure of varieties, with equal probability; $\mathcal{C}_{it}$ is the set of recipes that apply to variety $i$ at date $t$. Each recipe has a productivity that

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 $\sigma > 1$ | (14) |
| Variety $i$ output | $Y_{it} = Z_{Kit} \left( M_{it}^{-\frac{1}{\rho}} \sum_{j=1}^{M_{it}} x_{ijt}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}$ with $\rho > 1$ | (15) |
| Production of ingredients | $x_{ijt} = L_{ijt}$ | (16) |
| Best recipe | $Z_{Kit} = \max_{c \in \mathcal{C}_{it}} z_c$ | (17) |
| Weibull distribution of $z_c$ | $z_c \sim F(x) = 1 - e^{-x^{\beta}}$ | (18) |
| Number of ingredients evaluated | $\dot{N}_t = \alpha R_t^{\lambda} N_t^{\phi}, \quad \phi < 1$ | (19) |
| Cookbook (Poisson parameter) | $K_t = 2^{N_t}$ | (20) |
| Resource constraint: workers | $L_{it} = \sum_{j=1}^{M_i} L_{ijt}$ and $\int_0^1 L_{it} di = L_{yt}$ | (21) |
| Resource constraint: R&D | $R_t + L_{yt} = L_t$ | (22) |
| Population growth (exogenous) | $L_t = L_0 e^{g_L t}$ | (23) |

is i.i.d. with $z \sim F(z)$. For now, we assume the draws are from a Weibull distribution; in the next section, we will explain how this generalizes. One way to think about the randomness of the Poisson process versus the combinatorics associated with $2^N$ is that occasionally a recipe can apply to more than one variety or can be completely useless, and that is the randomness that allows the cookbook to contain more or fewer than $2^{N_t}$ recipes precisely at date $t$.

The Poisson parameter governing the evolution of recipes in the cookbook follows a *combinatorial growth process*, as defined earlier. That is, $K_t = 2^{N_t}$, where $N_t$ will (eventually) grow at a constant exponential rate. We generalize it slightly to incorporate two possible spillovers. 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 $\lambda$ allows for "stepping on toes" effects such as duplication, for example if $\lambda < 1$. The parameter $\phi$ 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 ($\phi < 0$).

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. This insight is captured in the current framework through the processing of ingredients. Imagine ingredients are ordered in such a way that the recipes for the telephone and radio get evaluated before the recipe for the smartphone. In that sense, the framework we've laid out incorporates the notion that the internet and television could not have been discovered before electricity.

### 3.3 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 (15) and (16) 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 (15) gives

$$Y_{it} = Z_{Kit}L_{it}. \tag{24}$$

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

$$\max_{\{L_{it}\}} Y_t = \left(\int_0^1 (Z_{Kit}L_{it})^{\frac{\sigma-1}{\sigma}}\, di\right)^{\frac{\sigma}{\sigma-1}} \tag{25}$$

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

$$Y_t = Z_{Kt}(1-\bar{s})L_t \;\text{ where} \tag{26}$$

$$Z_{Kt} = \left(\int_0^1 Z_{Kit}^{\sigma-1} di\right)^{\frac{1}{\sigma-1}} \tag{27}$$

Turning to the research side of the model,

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

This stable differential equation implies a constant asymptotic growth rate for $N$. 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

$$g_N \equiv \lim_{t\to\infty} \frac{\dot{N}_t}{N_t} = \frac{\lambda g_L}{1-\phi}. \tag{28}$$

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

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

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

Now we can integrate across the different sectors — and change the variable of integration to $\varepsilon$ — to get aggregate productivity. Now, however, we have to recall that a fraction $e^{-K}$ of our sectors will not have received any draws from the Poisson process and we assume their productivity defaults to zero. Therefore,

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

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

$$\boxed{g_y = g_Z = \frac{g_N}{\beta} = \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}.} \tag{29}$$

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 $\beta$, 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 $\beta$ 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, 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. This logic leads to a tighter intuition. Because $Z_K$ is a random variable, $\bar{F}(Z_K)$ is also a random variable. Importantly, recall that $\bar{F}(x)$ is uniformly distributed on $(0,1)$ when $x$ is a continuously-distributed random variable, 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 is an example of the result that $K$ times the minimum of $K$ draws from a $U(0,1)$ is asymptotically distributed as an exponential. This result is well-known in statistics and is just one special case of the extreme value theorem.[3] What is novel here is that the special case of $K\bar{F}(Z_K)$ is of particular interest: the fact that this random variable is asymptotically stationary has broad implications for how the max $Z_K$ increases with $K$.

### 4.2 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 the next, 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 3 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, $\eta(x) \equiv -\frac{d \log \bar{F}(x)}{d \log x}$. Then*

$$\lim_{t\to\infty} \frac{\dot{Z}_{Kt}}{Z_{Kt}} = \frac{g_N}{\alpha} \tag{30}$$

*if and only if*

$$\lim_{x\to\infty} \frac{\eta(x)}{x^{\alpha}} = \text{Constant} > 0 \tag{31}$$

[3] In particular, it leads to the third type of extreme value distribution, the Weibull, of which the exponential distribution is a special case.

*for some* $\alpha > 0$.

*Proof.* See Appendix A.3. QED

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.[4]

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 $\bar{F}(x) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$. 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 $g_N/2$.[5]

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 $\beta$ 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 $\beta$.

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 the next section.

[4] Klenow and Willis (2016) consider demand functions with this property. It would be interesting to see if such a demand function might be microfounded when taste heterogeneity has a Weibull distribution or a normal distribution.

[5] 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'Hopital'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.)

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 to 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.

### 4.3 Scaling and Growth for Other Distributions

The previous subsection characterized the class of distributions for which combinatorial growth in draws leads to exponential growth in the extreme value. 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

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

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.[6]

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 $x^\beta$ in the Weibull case. It is well known (and easy to show using Theorem 3 in Appendix A.1) that the Gompertz distribution is in the Gumbel domain of attraction. 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 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"

[6]To see another interesting application of this fact, suppose $\log x$ 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.

to the exponential distribution, in the sense that $\bar{F}(x)/\bar{G}(x) \to \text{Constant}$:

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

That is, for $x$ large, $e^{-e^{-x}} \approx 1 - e^{-x}$, so the Gumbel has an exponential upper tail. For this reason, the max grows directly with $\log K$, just like the exponential.

**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 $y$ 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):

$$\begin{aligned} K\bar{F}(Z_K) &= \varepsilon + o_p(1) \\ \Rightarrow \quad \frac{K}{(\log Z_K)^{\alpha}} &= \varepsilon + o_p(1) \\ \Rightarrow \quad \frac{\log Z_K}{K^{1/\alpha}} &= \left(\frac{1}{\varepsilon + o_p(1)}\right)^{1/\alpha} \end{aligned} \tag{32}$$

Next, because $\varepsilon$ is distributed as exponential with mean one, $\varepsilon^{-1/\alpha}$ is a Fréchet random variable with parameter $\alpha$.[7] Using this fact in equation (32) gives

$$\frac{\log Z_K}{K^{1/\alpha}} \overset{a}{\sim} \text{Fréchet}(\alpha) \tag{33}$$

and therefore

$$\log Z_K = K^{1/\alpha}(\tilde{\varepsilon} + o_p(1)) \tag{34}$$

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

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 $Z_K$.[8] In other words, if our productivity draws are log-Pareto distributed with the Pareto parameter equal to one, 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 (32), 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 (33), 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

[7]Since $\varepsilon$ has an exponential distribution with a mean 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

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

[8]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.

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: 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. 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 $g_N$.

observe; see Axtell (2001) and Luttmer (2010). In addition, if $K$ rises linearly, then the variance of log productivity would increase over time.[9] 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 $\tilde{\varepsilon}$ exist; the tail of the distribution is 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.

**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 Kortum (1997), an exponentially-growing number of draws from any distribution

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

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.

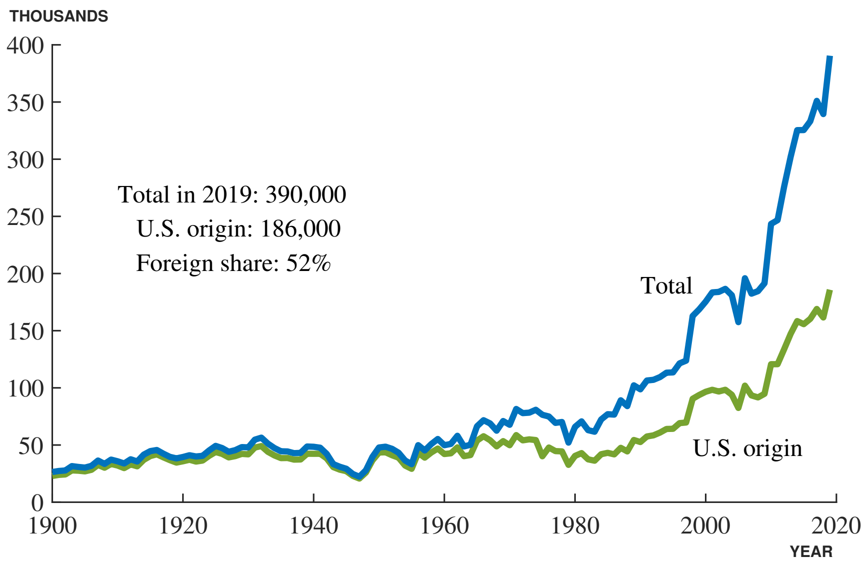
## 5. Evidence

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 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 a constant flow of new ideas can generate exponential growth.

However, even in the 1990s, the validity of this interpretation was unclear. 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. 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 8-fold 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; 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, the rise in patents may be driven by a few technologies, perhaps

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




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

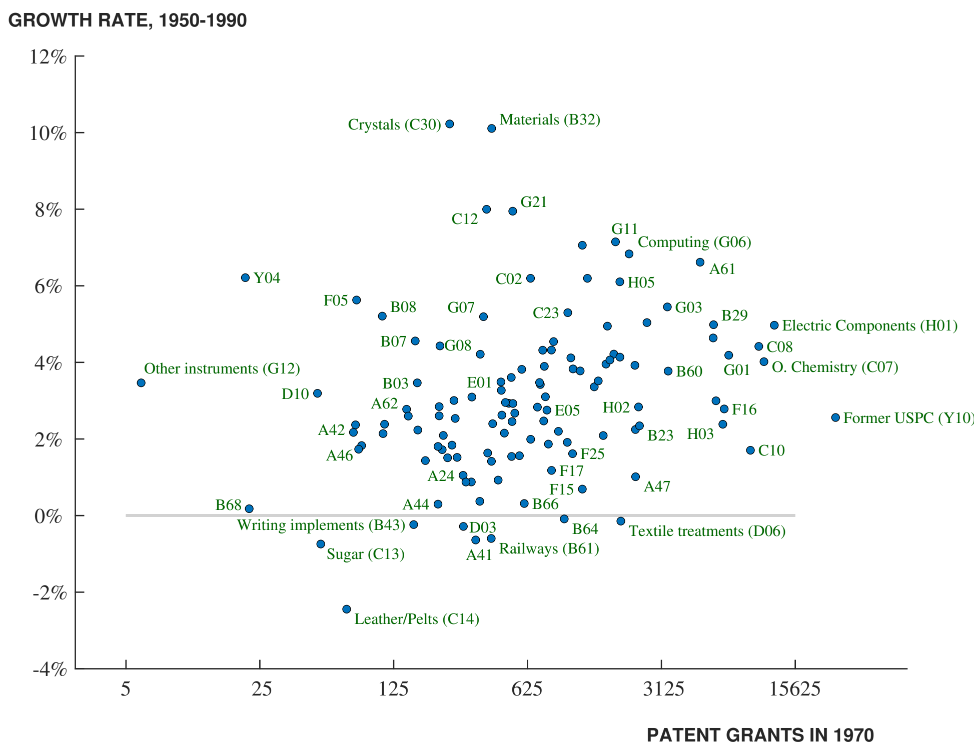
reflecting the rise of software and information technology.

Figure 2 shows the average annual growth rate of patents granted by the USPTO for 129 technology classes over the period 1950 to 1990, i.e. before the explosion of patenting associated with legal changes. Only 8 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, and the average growth rate rises slightly to 4.4% per year (though legal changes make this rise in the growth rate hard to interpret).

Alternatively, we can also consider a different measure of innovation: academic publications. Figure 3 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.

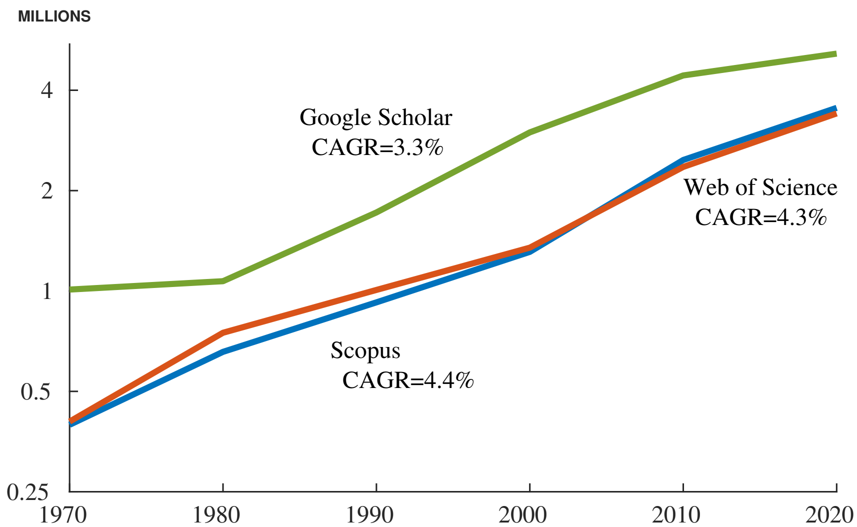
While none of these measures is perfect (and indeed, one drawback of the inno-

Figure 2: U.S. Patent Growth by Technology Class, 1950–1990




Note: The vertical axis shows the average annual growth rate of patents granted by the USPTO 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.

Figure 3: Annual Academic Publication Counts, 1970–2020




Note: Author's calculations using Scopus (2021), Web of Science (2021), and Google Scholar (2021), accessed June 30, 2021. CAGR denotes the cumulative average annual growth rate over the period.

vation literature is that we do not have great measures of innovation), they suggest that valuable new ideas are well-characterized as growing over time rather than being constant. 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 valuable new idea in the model. We follow Kortum (1997) in defining valuable new ideas as 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 $1/K_t$. In fact, this insight links very nicely with our main extreme value result. First, recall that the main result of Theorem 1 can be written as

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

Rearranging implies

$$\bar{F}(Z_K) = \frac{1}{K}(\varepsilon + o_p(1)). \tag{35}$$

In words, after $K$ draws, the probability that the next draw exceeds the max is approximately $1/K$. This is a nice connection between Theorem 1 and the theory of record breaking. The difference with the exact $1/K$ intuition given at the start of this paragraph is that now $Z_K$ is a random variable, but the spirit is the same.

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 simply $\frac{\dot{K}_t}{K_t}$. 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 $\phi = 0$).

$$\begin{aligned} K_t &= 2^{N_t} \\ \Rightarrow \frac{\dot{K}_t}{K_t} &= \log 2 \cdot \dot{N}_t \\ &= \log 2 \cdot \alpha R_t \\ &= \log 2 \cdot \alpha \bar{s} L_0 e^{g_L t} \end{aligned} \tag{36}$$

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.[10]

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 data on patents and publications, 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

[10] Kogan, Papanikolaou, Seru and Stoffman (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 $\phi > 0$.

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, $\alpha$ 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 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 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 (36) above, so each researcher would need to devote time to a constant number of new ideas, which seems reasonable.

## 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 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.[11] Their Corollary 2 suggests that broader results are possible with different growth rates for the number of new goods, and the present paper can be viewed as exploring those broader results.

Another paper that exploits combinations is Agrawal, McHale and Oettl (2019). They explore the effect of combininations on the idea production function and assume the elasticity of new ideas with respect to combinations declines to zero in order to prevent explosive growth.

**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 in this paper to follow the lead of Weitzman (1998) 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 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.

**Correlation.** A 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 but is beyond the scope of the present paper. Most of the results in the extreme value literature, for example, consider the 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

[11] 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.

result is likely to generalize to cases with correlation.

**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.

**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. 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 characterize the way in which the max $Z_K$ increases for any continuous distribution $\bar{F}(x)$ and any time path of (large) $K$.

# A. Appendix

## A.1 Extreme Value Theory

This appendix section provides a brief discussion of the standard Extreme Value Theorem and how it relates to results derived using Theorem 1 in the main text.

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.[12]

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 3.** *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*

$$\lim_{K\to\infty} \Pr\left[\frac{Z_K - b_K}{a_K} \leq x\right] = e^{-e^{-x}}. \tag{37}$$

*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 $a_K = KU'(K) = 1/(KF'(b_K))$.*

[12]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!

*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 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,

$$\frac{Z_K - b_K}{a_K} \overset{a}{\sim} \text{Gumbel} \tag{38}$$

Now we can show how this standard EVT result relates to the results derived in the paper. Letting $\varepsilon$ be a random variable from a standard Gumbel distribution, equation (38) is equivalent to

$$Z_K = b_K + a_K\varepsilon + o_p(a_K). \tag{39}$$

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).[13] Therefore, we have the result that

$$\boxed{\frac{Z_K}{b_K} \xrightarrow{p} 1.} \tag{40}$$

This is a special case of Theorem 4.1.1 in Galambos (1978) (his theorem further allows for dependence rather than assuming the draws are i.i.d.).

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, confirming the results derived in the main text for distributions in the Gumbel domain of attraction.

[13] 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.2 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 $\Pr\left[\, P = p \,\right] = \frac{e^{-K}K^p}{p!}$ so that $\Pr\left[\, P = 0 \,\right] = e^{-K}$ and $\Pr\left[\, P > 0 \,\right] = 1 - e^{-K}$. Therefore for $0 \leq m < K$

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

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

## A.3 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.

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 $-\frac{d\log\bar{F}(x)}{d\log x}\sim x^{\alpha}$, then $-\frac{d\log\bar{F}^{-1}(y)}{d\log y}\sim[\bar{F}^{-1}(y)]^{-\alpha}$ and vice versa since $y=\bar{F}(x)$. QED

**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 $0\leq\varepsilon<K$.

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 productivites. Changing the variable of integration from $i$ to $\varepsilon$ to take advantage of the continuum of varieties and recalling that the fraction $e^{-K_t}$ of sectors have zero Poisson draws and therefore zero productivity:

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

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've used the fact that $e^{-K_t}$ goes to zero to eliminate the first term.

**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 $-\frac{d\log \bar{F}^{-1}(\varepsilon/K)}{d\log(\varepsilon/K_t)} \sim \bar{F}^{-1}(\varepsilon/K_t)^{-\alpha}$ 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 $\psi$ 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 $\varepsilon$ and using 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 $\infty/\infty$ as $t \to \infty$, so we apply L'Hopital's rule to get the limit:

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

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

[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.

**If:** Now return to equation (41) for the "if" direction: if $\lim \frac{\dot{Z}_{Kt}}{Z_{Kt}} = g_N/\alpha$, then $\eta(x)$ is asymptotically a power function with exponent $\alpha$. 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 the value one 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 $\frac{\dot{K}_t}{K_t} = \dot{N} \log 2$ 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 $\varepsilon$. Substituting into the previous expression gives

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

since $\frac{-\log y}{\log K} \to 1$ for a fixed $\varepsilon$.

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 "$\sim$" version of this equation (being sure to keep $\alpha$ since the convergence is to 1 rather than to any constant) and then integrate:

$$\begin{aligned}
\frac{d\log \bar{F}^{-1}(y)}{d(-\log y)} &\sim \frac{1}{\alpha}\cdot\frac{1}{(-\log y)}\\
\Rightarrow\ d\log \bar{F}^{-1}(y) &\sim \frac{1}{\alpha}\cdot\frac{d(-\log y)}{(-\log y)}\\
\Rightarrow\ \int d\log \bar{F}^{-1}(y) &\sim \frac{1}{\alpha}\cdot\int\frac{d(-\log y)}{(-\log y)}\\
\Rightarrow\ \log \bar{F}^{-1}(y) &\sim \text{ Constant} + \frac{1}{\alpha}\log(-\log y)
\end{aligned}$$

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

$$\begin{aligned}
\Rightarrow\ -\log y &\sim x^{\alpha}\\
\Rightarrow\ -\log \bar{F}(x) &\sim x^{\alpha}\\
\Rightarrow\ -\frac{d\log \bar{F}(x)}{dx} &\sim \alpha x^{\alpha-1}\\
\Rightarrow\ -\frac{d\log \bar{F}(x)}{d\log x} &\sim x^{\alpha}
\end{aligned}$$

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

QED

## References

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

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

Agrawal, Ajay, John McHale, and Alexander Oettl, "Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019, pp. 149–174.

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

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

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

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.

Google Scholar, 2021. https://scholar.google.com, accessed June 30, 2021.

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.

Klenow, Peter J. and Jonathan L. Willis, "Real Rigidities and Nominal Price Changes," *Economica*, 2016, *83* (331), 443–472.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth," *The Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

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.

— , "Models of Growth and Firm Heterogeneity," *Annual Review Economics*, 2010, *2* (1), 547–576.

— , "Four Models of Knowledge Diffusion and Growth," Technical Report 2015. Federal Reserve Bank of Minneapolis working paper No. 724.

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.

Scopus, 2021. https://www.scopus.com/, accessed June 30, 2021.

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.

Web of Science, 2021. https://apps.webofknowledge.com, accessed June 30, 2021.

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