## EXTREME VALUE THEORY WITH HETEROGENEOUS AGENTS

### SEPHORAH MANGIN

Research School of Economics, Australian National University

Extreme value processes feature in any economic model in which agents receive a number of draws from some distribution and we examine the behavior of the maximum in the limit as the number of draws becomes large. This paper asks: Do the average outcomes of such processes change when different agents receive a different number of draws? To answer this, we allow the number of draws an agent receives from the underlying distribution (e.g. of productivities, ideas, or utility shocks) to be given by a *search technology*, which reflects heterogeneity in the expected number of draws across different types of agents. We derive a new class of extreme value distributions that generalize the three standard distributions (Fréchet, Gumbel, Weibull) by incorporating heterogeneity across agent types. We generalize a result from Gabaix, Laibson, Li, Li, Resnick, and de Vries (2016) regarding extreme value outcomes and consider applications to aggregate productivity, markups, and social networks.

KEYWORDS: Extreme value distribution, Fréchet distribution, Gumbel distribution, Mixed Poisson distribution, Heterogeneous agents, Productivity dispersion, Markups.

### 1. INTRODUCTION

EXTREME VALUE PROCESSES are widespread in economics. There is a large literature in which important economic outcomes such as output, productivity, growth, or markups – either at the firm level or the aggregate level – are determined by an extreme value process. Generally, such extreme value processes involve economic agents of some kind (e.g. firms, researchers, or consumers) receiving a number of draws from an underlying distribution of values (e.g. productivities, ideas, or utility shocks). We are typically interested in studying the distribution of the *maximum* in the limit as the number of draws becomes large.

It is well known that the limiting distribution of the (normalized) maximum must be one of three standard extreme value distributions: Fréchet, Gumbel, or Weibull. For example, if the

© 2025 The Authors. Econometrica published by John Wiley & Sons Ltd on behalf of The Econometric Society. Sephorah Mangin is the corresponding author on this paper. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Sephorah Mangin: sjmangin@gmail.com

I thank three anonymous referees for insightful suggestions that have significantly improved the paper. I thank my discussants, Jim Albrecht and Xiaoming Cai, as well as Brendan Beare, Saki Bigio, Paco Buera, Arthur Campbell, Begona Dominguez, Simon Grant, Tim Higgins, Michael Howes, Markus Kuba, Philipp Kircher, David Lagakos, Ruitian Lang, Hongyi Li, Daniel Luo, George Mailath, Guido Menzio, Ezra Oberfield, Guillaume Rocheteau, Shouyong Shi, Jeffrey Sun, Vitalii Tubdenov, Ludo Visschers, Pierre-Olivier Weill, Ronald Wolthoff, and Xinmei Yang for useful comments. I thank participants at the Barcelona Summer Forum, the D.C. Search and Matching Workshop, the Asia-Pacific Search and Matching Workshop, Tokyo, the West Coast Search and Matching Workshop, Vancouver, the Australian National University, the Bank of Canada, Cornell University, Deakin University, Queens University, UCLA, UC Santa Barbara, the University of Melbourne, the University of Sydney, the University of Toronto, UNSW, and the University of Queensland for useful feedback. This paper was originally submitted in 2023 as Becker and Mangin, "The Effect of Search Frictions on Extreme Outcomes." At the 2025 resubmission, Louis Becker requested to have his name removed as an author of the paper. I thank Louis Becker for his contribution and acknowledge the Australian National University for funding his position as a research assistant at an early stage of this project.

underlying distribution is Pareto, we obtain the Fréchet extreme value distribution.<sup>1</sup> Variations of the Pareto-Fréchet extreme value process are used in a large class of models in macroeconomics, growth, and trade, including papers by Kortum (1997), Eaton and Kortum (1999), Jones (2005), Lucas (2009), Mangin (2017), Buera and Lucas (2018), Oberfield (2018), Buera and Oberfield (2020), and Jones (2023).

In this paper, we provide a framework that generalizes such extreme value processes. Instead of assuming that all agents receive the same number of draws, we allow different agents to receive a different number of draws. We then examine the distribution of the (normalized) maximum in the limit as the *average* number of draws across agents becomes large.

To do this, we assume the number of draws an agent receives is itself a random variable that is given by a discrete probability distribution called the *search technology*. The search technology has two possible interpretations. First, it can be interpreted as reflecting randomness due to search or other frictions. Second, it can be interpreted as resulting from underlying heterogeneity across different *types* of agents. In this way, the results in this paper enable us to do extreme value theory in environments featuring heterogeneous agents.

The class of search technologies we consider is the class of mixed Poisson distributions. This class includes a wide range of discrete distributions, such as the negative binomial family, which nests both the geometric distribution and the Poisson distribution. Mixed Poisson distributions are particularly useful for modelling heterogeneity across different types of agents. A mixed Poisson random variable is a Poisson random variable where the Poisson parameter is itself a random variable that is distributed according to a "mixing distribution". For some applications, the distribution of the Poisson parameter can be interpreted as the *distribution of types*, where the "type" of an agent determines their expected number of draws.

The class of mixed Poisson distributions was introduced into economics by Hausman et al. (1984) as a way to model count data featuring overdispersion. In particular, the authors use this class to model heterogeneity across firms in R&D intensity.<sup>2</sup> In the search literature, mixed Poisson search technologies (or "meeting technologies") were first introduced by Cai et al. (2025), which microfounds this class of search technologies and uses it to model heterogeneity across locations in a spatial search model. Cai et al. (2025) proves that any search technology in an existing class of search technologies called "invariant" in Lester et al. (2015) can be represented as a mixed Poisson distribution. Therefore, all of our results apply to the class of invariant search technologies.

This paper provides two main results. First, we provide a general result regarding the extreme value distribution when the number of draws from the underlying distribution is random. Second, we provide a general result regarding the outcomes of extreme value processes, e.g. functions of maxima, when the number of draws is random. Our first main result extends a standard result in extreme value theory, the well-known Fisher-Tippett-Gnedenko extreme value theorem. Our second main result generalizes a result in Gabaix et al. (2016) regarding extreme value outcomes when the number of draws is fixed rather than random. We provide a formal proof of this general result which holds for any mixed Poisson search technology. We also provide a simple heuristic proof that uses an existing result in Mangin (2024) regarding extreme outcomes for the Poisson special case.

Intuitively, we might expect that, in the limit as the average number of draws becomes large, the effect of the search technology would disappear and the distribution of the maximum would

<sup>&</sup>lt;sup>1</sup>The Pareto distribution is all-pervasive in economics. See Gabaix (1999, 2009, 2016), Luttmer (2007), Benhabib and Bisin (2018), Jones and Kim (2018), Martellini and Menzio (2020), and Beare and Toda (2022).

<sup>&</sup>lt;sup>2</sup>More recently, mixed Poisson distributions have been used by Campbell et al. (2024) to model heterogeneity in a random network.

eventually behave in a standard way. Surprisingly, however, we find that the search technology still matters because there may be asymptotic dispersion in the number of draws even in this limit, which leads to greater dispersion in the distribution of the maximum. The search technology can therefore affect the asymptotic behavior of the maximum, or any functions of the maximum, and it may thereby affect important economic outcomes such as aggregate productivity or markups.

We find that the extreme value distribution need not be any of the three standard types (Fréchet, Gumbel, and Weibull). The form of the extreme value distribution depends not only on the underlying distribution and its tail index, but also on the search technology. For example, if the distribution of productivities is Pareto, the extreme value distribution is not necessarily Fréchet. We also find that the expected value of the maximum behaves asymptotically in a standard way except that it is scaled by a new term capturing the effect of the search technology. This scaling factor depends on both the tail index of the underlying distribution and on the search technology.

The effect of heterogeneity in the number of draws on extreme value outcomes is driven by the *asymptotic dispersion* of the search technology. The asymptotic dispersion can be interpreted as the residual dispersion that remains in the limit as the expected number of draws becomes large. Restricting attention to the mixed Poisson class of search technologies delivers significant tractability because the asymptotic dispersion of the search technology is simply equal to the dispersion of the type distribution. We find that extreme value outcomes may be either increasing or decreasing in a mean-preserving spread of the distribution of types. That is, greater asymptotic dispersion of the search technology – through greater dispersion in the expected number of draws across different types of agents – can either increase or decrease extreme value outcomes.

Our results show that the Poisson search technology is a unique special case where the search technology has no effect on extreme outcomes. Within the mixed Poisson class we consider, the extreme value distribution takes the standard form if and only if the search technology is Poisson. This is because the asymptotic dispersion of the Poisson search technology is zero, which explains why it delivers the standard extreme value results in the existing literature. The fact that the Poisson distribution can deliver the standard results in extreme value theory was first shown by Kortum (1997) in economics.<sup>3</sup> Variants of this Poisson approach were later used in Jones (2005), Mangin (2017), Mangin and Sedláček (2018), Oberfield (2018), Boehm and Oberfield (2023), and Jones (2023).

We consider an application of our results to aggregate productivity and the cross-sectional distribution of firm productivity. Firms are heterogeneous with respect to R&D intensity. Higher (lower) R&D intensity results in more (fewer) new ideas on average. We find that greater heterogeneity in R&D intensity *decreases* aggregate productivity if the underlying distribution is fat-tailed. Intuitively, this is because there are diminishing marginal returns to the number of ideas at a single firm. An additional idea is less valuable to a firm that already has a large stock of ideas, so aggregate productivity is maximized when there is less heterogeneity in R&D intensity across firms.

Next, we apply our results to the behavior of markups in a discrete choice model with random utility shocks. In this application, the search technology reflects the fact that consumers are heterogeneous with respect to their search intensities. Some consumers search more (less) intensely and find more (fewer) firms on average. We find that greater heterogeneity in consumers' search intensity *decreases* the average markup across consumers if the tail index of the

<sup>&</sup>lt;sup>3</sup>The basic idea was known in both statistics and in the "Peaks Over Threshold" model developed by hydrologists, which models the flood arrival rate as Poisson and the distribution of flood magnitudes as generalized Pareto and obtains an extreme value distribution. See Smith (1984).

distribution of utility shocks is positive (i.e. fat-tailed distributions) but *increases* the average markup if it is negative (e.g. uniform distribution). As we discuss, this is because the expected markup for a single consumer is asymptotically concave (convex) in the number of firms the consumer meets if the tail index is positive (negative), as shown in Gabaix et al. (2016).

Finally, we present an application to peer effects in a random social network of students. In this application, the search technology is the *degree distribution* of the social network and reflects the fact that friends are heterogeneous with respect to their popularity, i.e. their expected number of friends. We consider the effect of the degree distribution on the average student outcome in the limit as the network becomes dense, i.e. as the average number of friends becomes large. We find that greater heterogeneity in student popularity decreases the average outcome across students. Intuitively, this is because there are diminishing marginal returns to the number of friends of an individual student. An additional friend is more valuable to a less popular student compared to a more popular student, so the average outcome is maximized when the heterogeneity in popularity across students is lower.

Proofs of all results and some additional material can be found in the Appendix.

### 2. PRELIMINARIES

Suppose that agents receive a number of draws from some distribution of values. We assume that the distribution of values has cdf G and it satisfies Assumption 1. We call the distribution G the *underlying distribution*. Depending on the specific application, it may be a distribution of productivities, ideas, costs, valuations, or utility shocks.

ASSUMPTION 1: The distribution of values x has a twice-differentiable cdf G with pdf g = G' > 0, and support  $[\underline{x}, \overline{x}] \subseteq \mathbb{R}$  where  $\underline{x}, \overline{x} \in \mathbb{R} \cup \{\pm \infty\}$ , and  $\int_x^{\overline{x}} |xg(x)| dx$  is finite.<sup>4</sup>

We assume the number of draws is itself a random variable that has a discrete probability distribution with probability mass function  $P_n$  and support  $\mathbb{N}$ . For any  $n \in \mathbb{N}$ , the probability there are n draws is given by  $P_n(\theta)$  where  $\theta \in \mathbb{R}_+$  is the average number of draws across all agents. We denote by  $N(\theta)$  the random variable with distribution  $P_n$  and mean  $\theta$ . We refer to the distribution  $P_n$  as the *search technology*.

### 2.1. Mixed Poisson Class of Search Technologies

We restrict attention to search technologies that can be represented by mixed Poisson distributions. A *mixed Poisson* random variable is a Poisson random variable where the Poisson parameter is itself a random variable. In particular, we assume that the Poisson parameter is equal to a constant  $\theta \in \mathbb{R}_+$  multiplied by a random variable X, which is given by a continuous distribution F called the "mixing distribution", which we assume has a finite mean.<sup>5</sup>

We set  $\mathbb{E}_F[X] = 1$  to ensure the average number of draws across agents is  $\theta$ .<sup>6</sup>

<sup>&</sup>lt;sup>4</sup>More precisely, the support of G is an interval that may be either  $[\underline{x}, \overline{x}], (\underline{x}, \overline{x}], [\underline{x}, \overline{x}), \text{ or } (\underline{x}, \overline{x}).$ 

<sup>&</sup>lt;sup>5</sup>We focus on continuous mixing distributions F here, but some of our results could potentially be extended to discrete mixing distributions F. In particular, if there are  $K \in \mathbb{N}$  types  $(\tau_1, \tau_2, \ldots, \tau_K)$ , then (1) is replaced by  $P_n(\theta) = \sum_{j=1}^K \omega_j \frac{(\theta \tau_j)^n e^{-\theta \tau_j}}{n!}$  where  $\omega_j \equiv \Pr(X = \tau_j)$ . We leave this as an extension for future work.

<sup>&</sup>lt;sup>6</sup>The assumption that  $\mathbb{E}_{F}[X] = 1$  is not essential for our results. All of our main results still hold if  $\mathbb{E}_{F}[X] = \mu \in \mathbb{R}_{+}$ , but in this case the mean of the distribution  $P_{n}$  is  $\theta\mu$ .

ASSUMPTION 2—Mixed Poisson Search Technology: The search technology  $P_n$  is mixed Poisson with mean  $\theta \in \mathbb{R}_+$ . That is, for all  $n \in \mathbb{N}$ , we have

$$P_n(\theta) = \int_0^\infty \frac{(\theta\tau)^n e^{-\theta\tau}}{n!} dF(\tau)$$
(1)

for a continuous distribution with cdf F and support  $[\underline{\tau}, \overline{\tau}] \subseteq \mathbb{R}_+$  where  $\overline{\tau} \in \mathbb{R}_+ \cup \{+\infty\}$ .<sup>7</sup>

There are two possible interpretations of the mixed Poisson search technology.

First, the search technology  $P_n$  (e.g. geometric or negative binomial) may be the primitive and it could simply represent randomness in the number of draws received by different agents due to search or other frictions. In this case, the mixed Poisson representation is simply a convenient way to describe the search technology that is mathematically equivalent, but the mixing distribution F has no specific interpretation.

Second, the distribution F (e.g. exponential or gamma) may be the primitive and it could represent heterogeneity across different types of agents. For this reason, we refer to F throughout as the *distribution of types*. For agents of type  $\tau$ , the number of draws they receive is a Poisson random variable with mean  $\tau\theta$ . Higher (lower) types receive more (fewer) draws on average, but the average number of draws across *all* agents is  $\theta$ .

For example, suppose there is a large number of researchers who are heterogeneous with respect to their R&D intensity, which governs the rate of finding new ideas. For any given level of R&D intensity  $\tau$ , the distribution of the number of ideas a researcher gets is Poisson with mean  $\tau\theta$  for some constant  $\theta$ . This means that a higher R&D intensity leads to a higher expected number of ideas, but the realized number of ideas is still random. Now assume the distribution of R&D intensities  $\tau$  across researchers is given by a cdf F with mean  $\mathbb{E}_F[X] = 1$ . The search technology (i.e. the distribution of the number of ideas across all researchers in the economy) is mixed Poisson with mean  $\theta$  and satisfies Assumption 2. Of course, if all researchers have the same R&D intensity  $\tau = 1$ , we recover the standard Poisson search technology.

### 2.2. Useful Properties of Mixed Poisson Distributions

Mixed Poisson distributions have some useful properties that are shared with the Poisson distribution. One useful property is that the probability generating function of  $P_n$  can be expressed in terms of the function  $P_0 : \mathbb{R}^+ \to [0, 1]$  where  $P_0(\theta)$  is the probability of receiving zero draws when the average number of draws is  $\theta$ .<sup>8</sup> This means the function  $P_0$  captures everything we need to know about the search technology.<sup>9</sup>

LEMMA 1: If the search technology  $P_n$  is mixed Poisson with mean  $\theta$ , then for  $y \in [0, 1]$ ,

$$\mathbb{E}_P[y^{N(\theta)}] = \sum_{n=0}^{\infty} P_n(\theta) y^n = P_0(\theta(1-y)).$$

<sup>&</sup>lt;sup>7</sup>More precisely, the support of F is an interval that may be either  $[\underline{\tau}, \overline{\tau}]$  if  $\overline{\tau} < \infty$  or  $[\underline{\tau}, \overline{\tau})$  if  $\overline{\tau} = \infty$ .

<sup>&</sup>lt;sup>8</sup>It is worth pointing out that (1) implies  $P_0(z) = M_F(-z)$  where  $M_F$  is the moment generating function of the distribution F. This means that in applications where we start with a type distribution F, the function  $P_0$  can be directly calculated if we know the moment generating function of F. Note that since z > 0, the moment-generating function need only be defined on the negative reals.

<sup>&</sup>lt;sup>9</sup>The property in Lemma 1 is a property of the class of meeting technologies called *invariant* in Lester et al. (2015). In Appendix E, we describe an equivalence between the class of mixed Poisson distributions and the class of search technologies called invariant in Lester et al. (2015). This equivalence result is established in Cai et al. (2025). Therefore, all of our results apply to any search technology that is "invariant" as defined in Lester et al. (2015). Note that invariance also implies *non-rivalry*, as defined in Eeckhout and Kircher (2010).

6

#### PROOF: See Appendix A.

Another useful property of this class is that the sum of two mixed Poisson random variables is itself a mixed Poisson random variable.<sup>10</sup> Under certain conditions, the distribution of the sum is a search technology of the same kind. For example, a finite sum of negative binomial random variables is itself negative binomial under certain conditions.<sup>11</sup>

In Appendix F, we show how the mixed Poisson assumption can be used in dynamic environments where agents can accumulate draws over time.

### 2.3. Asymptotic Dispersion of Search Technology

For any mixed Poisson distribution with  $\mathbb{E}_F[X] = 1$ , the squared coefficient of variation is a measure of dispersion given by

$$cv_P^2(\theta) = \frac{1}{\theta} + cv_F^2 \tag{2}$$

where  $cv_F^2$  is the squared coefficient of variation of the mixing distribution F.<sup>12</sup> This measure of dispersion has two components: the first reflects the dispersion of the Poisson distribution and the second captures the dispersion of the mixing distribution.

We are interested in what happens when the expected number of draws becomes large. A key object of interest will be the *asymptotic dispersion* of the search technology, which we define as  $cv_P^2 \equiv \lim_{\theta \to \infty} cv_P^2(\theta)$ . This captures the "residual" dispersion that remains in the limit as the expected number of draws becomes large. It is clear from expression (2) that we have  $\lim_{\theta \to \infty} cv_P^2(\theta) = cv_F^2$ . That is, the asymptotic dispersion of the search technology  $P_n$  is equal to the dispersion of the type distribution F.

LEMMA 2—Asymptotic Dispersion: If the search technology  $P_n$  is mixed Poisson with mean  $\theta$  and  $\mathbb{E}_F[X] = 1$ , the asymptotic dispersion  $cv_P^2$  of the search technology is equal to the dispersion  $cv_F^2$  of the type distribution F. That is,

$$cv_P^2 = \lim_{\theta \to \infty} cv_P^2(\theta) = cv_F^2.$$

For the Poisson distribution, there is no asymptotic dispersion because  $cv_F^2 = 0$ . In this special case, the distribution F is degenerate and the random variable  $N(\theta)$  behaves asymptotically like a constant (i.e. as though  $P_n$  is asymptotically degenerate). However, if the search technology  $P_n$  is not Poisson – or, equivalently, the distribution F is non-degenerate – then  $N(\theta)$  exhibits asymptotic dispersion. We will later see that asymptotic dispersion, i.e. the dispersion of the type distribution F, will be important for understanding the impact of the search technology on extreme outcomes.

<sup>&</sup>lt;sup>10</sup>In general, if the mixing distributions are  $F_1$  and  $F_2$ , the mean of the sum is the sum of the two means and the mixing distribution of the sum is the convolution of the two distributions  $F_1$  and  $F_2$ .

<sup>&</sup>lt;sup>11</sup>For example, the sum of two negative binomial random variables with parameters  $r_1$ ,  $r_2$  and means  $\theta_1$ ,  $\theta_2$  is itself negative binomial with parameter  $r_1 + r_2$  and mean  $\theta_1 + \theta_2$  if  $\theta_1/\theta_2 = r_1/r_2$ . This includes the special case where the two negative binomial random variables are the same, i.e.  $r_1 = r_2$  and  $\theta_1 = \theta_2$ . See Johnson et al. (2005).

<sup>&</sup>lt;sup>12</sup>The squared coefficient of variation is equal to the variance divided by the squared mean. For any mixed Poisson distribution with  $\mathbb{E}_F[X] = 1$ , the variance is  $\sigma_P^2(\theta) = \theta + \sigma_F^2 \theta^2$  where  $\sigma_F^2$  is the variance of the mixing distribution F. See Karlis and Xekalaki (2005) for further details.

#### 2.4. Distribution of the Maximum

Importantly, Lemma 1 enables us to derive a simple expression for the distribution of the maximum when the number of draws is a mixed Poisson random variable. This result may be independently useful in cases where the *exact* distribution for a finite number of draws is required, instead of the extreme value distribution that arises in the limit as the number of draws becomes large – which we derive in Section 4.

Define a random variable for the maximum of  $N(\theta)$  draws, where  $N(\theta)$  is a mixed Poisson random variable, by  $M_{N(\theta)} \equiv \max\{X_1, \ldots, X_{N(\theta)}\}^{13}$  Define the cdf of the distribution of the maximum by  $H_P(x;\theta) \equiv \Pr(M_{N(\theta)} \leq x)$ . It is well known that the distribution of the maximum of *n* draws from *G* has cdf  $G(x)^n$ . Weighting the conditional distribution  $G(x)^n$ (i.e. conditional on *n* draws) by the probability  $P_n(\theta)$  of realizing *n* draws,

$$H_P(x;\theta) = \sum_{n=0}^{\infty} P_n(\theta) G(x)^n.$$

We can now apply Lemma 1, setting y = G(x), to derive the following result.

LEMMA 3—Distribution of the Maximum: If the search technology  $P_n$  is mixed Poisson with mean  $\theta$ , the distribution of the maximum of  $N(\theta)$  draws from a distribution G is

$$H_P(x;\theta) = P_0(\theta(1 - G(x))). \tag{3}$$

We illustrate by providing some examples of mixed Poisson search technologies.

### 2.5. Examples

In this section, we present some important examples of mixed Poisson search technologies. Table II in Appendix E provides some additional examples.<sup>14</sup>

EXAMPLE 2.1—Poisson: If  $P_n$  is a Poisson search technology, then

$$P_n(\theta) = \frac{e^{-\theta}\theta^n}{n!}$$

and the function  $P_0$  is

$$P_0(z) = e^{-z}$$

Therefore, Lemma 3 tells us that the distribution of the maximum is

$$H_P(x;\theta) = e^{-\theta(1-G(x))}$$

We can represent the search technology  $P_n$  using (1) where the mixing distribution F is degenerate and has mean one. The asymptotic dispersion of  $P_n$  is  $cv_P^2 = cv_F^2 = 0$ .

EXAMPLE 2.2—Negative Binomial: If  $P_n$  is a negative binomial search technology,

$$P_n(\theta) = \binom{n+r-1}{n} \left(\frac{r}{r+\theta}\right)^r \left(\frac{\theta}{r+\theta}\right)^n$$

<sup>&</sup>lt;sup>13</sup>To simplify, we assume that  $M_{N(\theta)} = \underline{x}$  in the case where there are zero draws.

<sup>&</sup>lt;sup>14</sup>See also Karlis and Xekalaki (2005) for a detailed list of mixed Poisson distributions.

where  $r \in \mathbb{N} \setminus \{0\}$ , and the function  $P_0$  is

$$P_0(z) = \left(\frac{r}{r+z}\right)^r.$$
(4)

Therefore, by Lemma 3, the distribution of the maximum is

$$H_P(x;\theta) = \left(\frac{r}{r+\theta(1-G(x))}\right)^r$$

We can represent the negative binomial search technology  $P_n$  using (1) where the mixing distribution F is the gamma distribution with support  $\tau \in [0, \infty)$  and cdf  $F(\tau) = \frac{\gamma(r, r\tau)}{\Gamma(r)}$  where  $\mathbb{E}_F[X] = 1$ . The asymptotic dispersion of  $P_n$  is  $cv_P^2 = cv_F^2 = 1/r$ .

Intuitively, we can interpret the negative binomial random variable  $N(\theta)$  as counting the number n of successes before r failures, where the probability of success is given by  $\theta/(r+\theta)$  and the expected number of successes before r failures is  $\theta$ . In the limit as  $r \to \infty$ , the random variable  $N(\theta)$  counts the total number of successes as the probability of success becomes small, i.e.  $N(\theta)$  is a Poisson random variable. It is easy to verify directly from the  $P_0$  function (4) that, in the limit as  $r \to \infty$ , we obtain the Poisson search technology.<sup>15</sup>

EXAMPLE 2.3—Geometric: If  $P_n$  is a geometric search technology, then

$$P_n(\theta) = \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta}\right)^n$$

and the function  $P_0$  is

$$P_0(z) = \frac{1}{1+z}.$$

Therefore, by Lemma 3, the distribution of the maximum is

$$H_P(x;\theta) = \frac{1}{1+\theta(1-G(x))}.$$

The geometric search technology is a special case of the negative binomial family where r = 1. We can represent the search technology  $P_n$  using (1) where the mixing distribution F is the exponential distribution with support  $\tau \in [0, \infty)$  and cdf  $F(\tau) = 1 - e^{-\tau}$  where  $\mathbb{E}_F[X] = 1$ . The asymptotic dispersion of  $P_n$  is  $cv_P^2 = cv_F^2 = 1$ .

Figure 1 depicts the negative binomial family of search technologies for large  $\theta$ . In the limit as  $r \to \infty$  and  $P_n$  is Poisson, the random variable  $N(\theta)$  behaves asymptotically like a constant because  $cv_F^2 = 1/r \to 0$ . However, if the search technology  $P_n$  is not Poisson, then  $N(\theta)$  exhibits asymptotic dispersion. We will see that this asymptotic dispersion matters for both extreme value outcomes and the extreme value distribution.

<sup>15</sup>Note that 
$$\lim_{r \to \infty} \left(1 + \frac{z}{r}\right)^r = e^z$$
, so  $\lim_{r \to \infty} P_0(z) = \lim_{r \to \infty} \left(\frac{r}{r+z}\right)^r = \lim_{r \to \infty} \left(1 + \frac{z}{r}\right)^{-r} = e^{-z}$ .



FIGURE 1.—Search Technologies in the Negative Binomial Family ( $\theta = 100$ ).

#### 3. PREVIEW: PARETO EXAMPLE

In this section, we preview some of our results by providing a simple heuristic derivation of the extreme value distribution for a specific example where the underlying distribution is Pareto. This enables us to see immediately why the search technology matters for determining the shape of the extreme value distribution.

We first derive the standard result for a fixed number of draws, and then show how this generalizes to our environment where the number of draws is a random variable. Our results are closely connected to a result in Jones (2023) regarding the asymptotic behavior of the maximum. We provide details of this connection in Appendix G.

#### 3.1. Fixed Number of Draws

Let  $X_1, \ldots, X_n$  be i.i.d. random variables with distribution G where  $n \ge 1$ . Define a random variable for the maximum,  $M_n \equiv \max\{X_1, \ldots, X_n\}$ . It is well known that the cdf of the distribution of the maximum is given by  $H(x; n) \equiv \Pr(M_n \le x) = G(x)^n$ .

Suppose the underlying distribution is Pareto, i.e.  $G(x) = 1 - x^{-1/\gamma}$ . Given that  $M_n \to \infty$  as  $n \to \infty$ , we need to normalize the random variable  $M_n$ . In this example, we can "guess" and verify that the right normalization is to define a new random variable  $Z_n \equiv M_n/n^{\gamma}$ .

The fact that  $Pr(M_n \le x) = G(x)^n$  gives us

$$\Pr\left(M_n \le \left(\frac{y}{n}\right)^{-\gamma}\right) = \left(1 - \frac{y}{n}\right)^n.$$

As  $n \to \infty$ , the right-hand side converges to  $e^{-y}$ , so in the limit as  $n \to \infty$ , we have

$$\Pr(Z_n \le y^{-\gamma}) = e^{-y}.$$

Letting  $x = y^{-\gamma}$ , we have  $\Pr(Z_n \le x) = e^{-x^{-1/\gamma}}$  in the limit as  $n \to \infty$ . Therefore, the distribution of the normalized maximum is given by the following extreme value distribution:

$$H_{\gamma}(x) = e^{-x^{-1/\gamma}}.$$
(5)

As expected, the Pareto distribution delivers a Fréchet extreme value distribution.

### 3.2. Random Number of Draws

Consider the same environment except for one difference: the number of draws is a random variable  $N(\theta)$  with mean  $\theta$  and distribution  $P_n$ , which we assume is mixed Poisson. Define a random variable for the maximum,  $M_{N(\theta)} \equiv \max\{X_1, \ldots, X_{N(\theta)}\}$ . From expression (3), we have  $H_P(x;\theta) \equiv \Pr(M_{N(\theta)} \le x) = P_0(\theta(1 - G(x)))$ .

Suppose the underlying distribution is Pareto, i.e.  $G(x) = 1 - x^{-1/\gamma}$ . Given that  $M_{N(\theta)} \to \infty$  as  $\theta \to \infty$ , we need to normalize the random variable  $M_{N(\theta)}$ . We can again "guess" and verify that the right normalization is to define a new random variable  $Z_{N(\theta)} \equiv M_{N(\theta)}/\theta^{\gamma}$ . The fact that  $\Pr(M_{N(\theta)} \le x) = P_0(\theta(1 - G(x)))$  by Lemma 3 gives us<sup>16</sup>

$$\Pr(Z_{N(\theta)} \le y^{-\gamma}) = P_0(y)$$

Again letting  $x = y^{-\gamma}$ , we have  $\Pr(Z_{N(\theta)} \le x) = P_0(x^{-1/\gamma})$ . So, the distribution of the normalized maximum is given by the following extreme value distribution:

$$H_{\gamma,P}(x) = P_0(x^{-1/\gamma}).$$
 (6)

In the special case where the search technology  $P_n$  is Poisson, we have  $P_0(z) = e^{-z}$  and we therefore recover (5), the Fréchet extreme value distribution.

For example, if  $P_n$  is a negative binomial search technology and  $r \in \mathbb{N} \setminus \{0\}$ , then we have  $P_0(z) = \left(\frac{r}{r+z}\right)^r$  and the extreme value distribution is

$$H_{\gamma,P}(x) = \left(\frac{r}{r+x^{-1/\gamma}}\right)^r$$

This is a generalization of the Fréchet distribution. As  $r \to \infty$  and the negative binomial search technology converges to the Poisson, we recover the Fréchet.

Figure 2 provides an illustration that shows how the extreme value distribution varies with the search technology  $P_n$ . For the Poisson distribution, there is no asymptotic dispersion, so the extreme value distribution is the standard Fréchet distribution. However, as we move away from the Poisson, the additional dispersion arising from the search technology leads to a first-order stochastic dominance shift in the extreme value distribution. In Section 4, we prove that this first-order stochastic dominance result holds generally.

#### 4. EXTREME VALUE DISTRIBUTION

In this section, we provide a general result regarding the form of the extreme value distribution when the number of draws is random and the search technology is mixed Poisson.

We first present the standard result for the extreme value distribution when the number of draws is fixed, and then provide our novel result for a random number of draws.

<sup>&</sup>lt;sup>16</sup>Notice that this result does not require taking  $\theta \to \infty$ , as it does for the case where n is fixed.



FIGURE 2.—Extreme Value Distributions for Negative Binomial Family (G is Pareto and  $\gamma = 1/2$ ).

We assume throughout the remainder of the paper that G is well-behaved in the sense that it has a finite tail index  $\gamma \in \mathbb{R}$  where the *tail index* is defined as follows.

DEFINITION 1—Tail Index: The *tail index*  $\gamma \in \mathbb{R}$  of a distribution G is given by

$$\gamma \equiv \lim_{x \to \overline{x}} \frac{d}{dx} \left( \frac{1 - G(x)}{g(x)} \right).$$

The tail index is a measure of tail fatness, with a higher value of  $\gamma$  corresponding to fatter tails. It is critical for determining the type of the extreme value distribution.

#### 4.1. Fixed Number of Draws

Let  $X_1, \ldots, X_n$  be i.i.d. random variables with distribution G. Define the random variable  $M_n \equiv \max\{X_1, \ldots, X_n\}$  where  $n \ge 1$ . Given that we assume G is well-behaved with tail index  $\gamma$ , in the sense of Definition 1, classical extreme value theory tells us that there exist normalizing constants  $a_n$ ,  $b_n$  such that the sequence of normalized random variables  $Z_n \equiv a_n M_n + b_n$  converges in distribution as  $n \to \infty$  to the following:

$$H_{\gamma}(x) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}} & \text{if } \gamma \neq 0, \\ e^{-e^{-x}} & \text{if } \gamma = 0, \end{cases}$$
(7)

where  $H_{\gamma}(x) \equiv \lim_{n \to \infty} \Pr(Z_n \leq x)$ . We say that G is in the domain of attraction of the extreme value distribution  $H_{\gamma}$ .<sup>17</sup> This distribution must be one of only three types: Fréchet (if  $\gamma > 0$ ), reverse-Weibull (if  $\gamma < 0$ ), or Gumbel (if  $\gamma = 0$ ).<sup>18</sup>

### 4.2. Random Number of Draws

Now suppose  $N(\theta)$  is a discrete random variable with mean  $\theta$ , and define the random variable  $M_{N(\theta)} \equiv \max\{X_1, \ldots, X_{N(\theta)}\}$ . Theorem 1 tells us that there exist normalizing constants  $a_{\theta}$ ,  $b_{\theta}$  such that the sequence of normalized random variables  $Z_{N(\theta)} \equiv a_{\theta} M_{N(\theta)} + b_{\theta}$  converges in distribution as  $\theta \to \infty$  and the limiting distribution (i.e. the extreme value distribution) takes a form which depends on both the tail index  $\gamma$  and the search technology  $P_n$ .

THEOREM 1—Extreme Value Distribution: Suppose  $N(\theta)$  is a random variable with mixed Poisson distribution  $P_n$  and mean  $\theta$ . Let  $a_n$ ,  $b_n$  be constants such that the distribution of the normalized random variable  $Z_n \equiv a_n M_n + b_n$  converges as  $n \to \infty$  to the extreme value distribution  $H_{\gamma}(x) \equiv \lim_{n \to \infty} \Pr(Z_n \leq x)$  given by

$$H_{\gamma}(x) = e^{-v_{\gamma}(x)}$$

where  $v_{\gamma}(x) \equiv (1 + \gamma x)^{-1/\gamma}$  if  $\gamma \neq 0$  and  $v_{\gamma}(x) \equiv e^{-x}$  if  $\gamma = 0$ . The distribution of the normalized random variable  $Z_{N(\theta)} \equiv a_{\theta} M_{N(\theta)} + b_{\theta}$  converges as  $\theta \to \infty$  to the extreme value distribution  $H_{\gamma,P}(x) \equiv \lim_{\theta \to \infty} \Pr(Z_{N(\theta)} \leq x)$  given by

$$H_{\gamma,P}(x) = P_0(v_\gamma(x)). \tag{8}$$

Q.E.D.

PROOF: See Appendix B.1.

Theorem 1 is useful because it allows us to generalize the standard results in extreme value theory to incorporate heterogeneous agents – who may receive a different number of draws – into any economic model involving an extreme value process.

The form of the extreme value distribution (8) is the same as for the Pareto example (6)in Section 3. In general, the standard extreme value distribution  $H_{\gamma}$  first-order stochastically dominates  $H_{\gamma,P}$  for any search technology that is not Poisson, i.e.  $P_0(v_{\gamma}(x)) > e^{-v_{\gamma}(x)}$  for any x. As a result, the expected value of the distribution  $H_{\gamma,P}$  is always strictly lower than the expected value of the distribution  $H_{\gamma}$  if the search technology  $P_n$  is not Poisson.<sup>19</sup>

Corollary 1 summarizes this result. Intuitively, the first-order stochastic dominance arises because there are diminishing marginal returns to the number of draws for any given agent. This means that the expected value of the maximum draw across agents is decreasing in the degree of asymptotic dispersion in the number of draws, which is equal to the degree of heterogeneity across agents. From this perspective, greater dispersion in the number of draws can be viewed as a kind of "misallocation" in the number of draws across agents.

<sup>&</sup>lt;sup>17</sup>See, for example, Theorem 1.1.8 in de Haan and Ferreira (2006). Examples of such normalizing constants are  $a_n = 1/G^{-1}(1-1/n)$  and  $b_n = 0$  for fat-tailed distributions with  $\gamma > 0$ .

<sup>&</sup>lt;sup>18</sup>To see this, observe that  $H_{\gamma}(x) = e^{-e^{-x}}$  (Gumbel) for  $\gamma = 0$ ,  $H_{\gamma}((x-1)/\gamma) = e^{-x^{-1/\gamma}}$  (Fréchet) for  $\gamma > 0$ , and  $H_{\gamma}(-(x+1)/\gamma) = e^{-(-x)^{-1/\gamma}}$  (reverse-Weibull) for  $\gamma < 0$ . <sup>19</sup>Assumption 2 implies  $P_0(z) = \mathbb{E}_F[e^{-zX}]$ . Since  $e^{-zX}$  is a convex function of X for any  $z \in \mathbb{R}_+$ , Jensen's inequality implies  $\mathbb{E}_F[e^{-zX}] > e^{-z\mathbb{E}_F[X]} = e^{-z}$  and therefore  $P_0(z) > e^{-z}$  for any  $z \in \mathbb{R}_+$ .

COROLLARY 1—First-Order Stochastic Dominance: If a mixed Poisson search technology  $P_n$  is not Poisson, the standard extreme value distribution  $H_{\gamma}$  first-order stochastically dominates the extreme value distribution  $H_{\gamma,P}$ .

In the special case where  $P_n$  is Poisson and  $P_0(z) = e^{-z}$ , it is easy to see that we recover the standard form of the extreme value distribution, i.e.  $H_{\gamma,P}(x) = H_{\gamma}(x)$ . Conversely, if  $H_{\gamma,P}(x) = H_{\gamma}(x)$  and  $P_n$  is mixed Poisson, then it must be Poisson.<sup>20</sup>

COROLLARY 2—Uniqueness of Poisson: For any mixed Poisson search technology  $P_n$ , the extreme value distribution is standard, i.e.  $H_{\gamma,P}(x) = H_{\gamma}(x)$ , if and only if  $P_n$  is Poisson.

As we discussed, the reason behind the unique role of the Poisson distribution is the fact that the mixing distribution F is degenerate and there is no asymptotic dispersion. The random variable  $N(\theta)$  behaves asymptotically like a constant. However, if the search technology  $P_n$  is not Poisson then  $N(\theta)$  exhibits asymptotic dispersion and the extreme value distribution does not take any of the three standard forms.

Our class of extreme value distributions shares with the three standard extreme value distributions a convenient property. Distributions in this class are *max stable* in the following sense. If we take a distribution of the form (8) and use this as our underlying distribution G, and then apply the same mixed Poisson search technology  $P_n$ , the resulting extreme value distribution delivered by Theorem 1 is the same distribution G. That is, the extreme value distribution lies within its *own* domain of attraction, i.e. it is max stable.

THEOREM 2—Max Stability: Suppose the underlying distribution G is an extreme value distribution with cdf  $G(x) = P_0(v_{\gamma}(x))$  for some mixed Poisson search technology  $P_n$  and  $\gamma \in \mathbb{R}$ . The distribution G lies in its own domain of attraction, i.e.  $H_{\gamma,P} = G$ .

PROOF: See Appendix B.2.

Our next result regarding the tail index follows from Theorems 1 and 2.

COROLLARY 3—Inheritance of Tail Index: Suppose that  $H_{\gamma,P}(x) = P_0(v_{\gamma}(x))$  for some mixed Poisson search technology  $P_n$  and an underlying distribution G with tail index  $\gamma \in \mathbb{R}$ . If the distribution  $H_{\gamma,P}$  has tail index  $\gamma_H$ , then it inherits the same tail index, i.e.  $\gamma_H = \gamma$ .

PROOF: See Appendix B.3.

Corollary 3 says that the tail index  $\gamma$  of an underlying distribution G is inherited by the extreme value distribution  $H_{\gamma,P}$ . Therefore, regardless of the search technology  $P_n$ , the extreme value distribution always retains the same tail index (see Figure 2).

#### 4.3. Examples

We now present some examples of the extreme value distribution for different search technologies and a general underlying distribution G.

Q.E.D.

<sup>&</sup>lt;sup>20</sup>To see this, if  $P_n$  is mixed Poisson and  $H_{\gamma,P}(x) = H_{\gamma}(x)$ , then  $P_0(v_{\gamma}(x)) = e^{-v_{\gamma}(x)}$  by Theorem 1, so  $P_0(z) = e^{-z}$ . Thus the probability generating function of  $P_n$  is  $e^{-\theta(1-y)}$  by Lemma 1 and  $P_n$  must be Poisson because a distribution is uniquely determined by its probability generating function.

EXAMPLE 4.1—Poisson: If  $P_n$  is Poisson, the extreme value distribution is

$$H_{\gamma,P}(x) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}} & \text{if } \gamma \neq 0, \\ e^{-e^{-x}} & \text{if } \gamma = 0. \end{cases}$$

That is, the Poisson search technology delivers the standard extreme value distribution not only for the Pareto example in Section 3 but for *any* underlying distribution.

EXAMPLE 4.2—Negative Binomial: If  $P_n$  is a negative binomial search technology,

$$H_{\gamma,P}(x) = \begin{cases} \left(\frac{r}{r+(1+\gamma x)^{-1/\gamma}}\right)^r & \text{if } \gamma \neq 0, \\ \left(\frac{r}{r+e^{-x}}\right)^r & \text{if } \gamma = 0, \end{cases}$$

where  $r \in \mathbb{N} \setminus \{0\}$ . Clearly, this differs from the standard extreme value distribution above.<sup>21</sup>

### 5. EXTREME VALUE OUTCOMES

In this section, we provide a general result regarding the outcomes of extreme value processes when the number of draws is random and the search technology is mixed Poisson.

More precisely, we study the asymptotic behavior of the expected value of functions  $\zeta$  of the maximum  $M_{N(\theta)}$  where the number of draws  $N(\theta)$  is a random variable. Specifically, we take the expectation of  $\zeta(M_{N(\theta)})$  with regard to the distribution of the maximum,  $H_P(x;\theta)$ , and then consider the limit as the expected number of draws  $\theta$  goes to infinity. That is, we consider the asymptotic behavior of the *extreme value outcome*,  $\mathbb{E}_{H_P}[\zeta(M_{N(\theta)})]$ .

We first state the original result from Gabaix et al. (2016) regarding extreme value outcomes when the number of draws is fixed, and then present our result for a random number of draws. We require the following definition of *regular variation*.<sup>22</sup>

DEFINITION 2—Regular Variation: A function  $k : \mathbb{R}_+ \to \mathbb{R}$  is regularly varying at zero with index  $\rho$ , denoted  $k(t) \in RV_{\rho}^0$ , if and only if, for all a > 0, we have

$$\lim_{t \to 0} \frac{k(at)}{k(t)} = a^{\rho}$$

The following assumption regarding the function  $\zeta$  will be used in Theorem 3.

ASSUMPTION 3: The function  $\zeta : [\underline{x}, \overline{x}] \to \mathbb{R}_+$  is measurable and bounded on  $[\underline{x}, x]$  for any  $x \in (\underline{x}, \overline{x})$ , and  $\int_{\underline{x}}^{\overline{x}} |\zeta(x)g(x)| dx$  is finite.

In our statement of the following results, we adopt standard notation and we write  $h_1(\theta) \sim_{\theta \to \infty} h_2(\theta)$ , or simply  $h_1(\theta) \sim h_2(\theta)$ , if and only  $\lim_{\theta \to \infty} h_1(\theta)/h_2(\theta) = 1$ .

<sup>&</sup>lt;sup>21</sup>The extreme value distributions derived here are consistent with those derived in Section 3 for  $\gamma \neq 0$  except for the linear transformation  $1 + \gamma x$ . This allows the flexibility to accommodate  $v_{\gamma}(x) = e^{-x}$  in the limit as  $\gamma \to 0$  since we define  $v_{\gamma}(x) \equiv (1 + \gamma x)^{-1/\gamma}$  if  $\gamma \neq 0$ . This form is consistent with the generalized extreme value distribution (7) which incorporates all three standard types.

<sup>&</sup>lt;sup>22</sup>See Bingham et al. (1987) or Resnick (1987). Notice that  $k(t) \in RV_{\rho}^{0}$  is equivalent to  $\hat{k}(t) \equiv k(1/t) \in RV_{-\rho}^{\infty}$ , which we use whenever convenient.

#### 5.1. Fixed Number of Draws

Let  $X_1, \ldots, X_n$  be i.i.d. random variables with distribution G where  $n \ge 1$  and define  $M_n \equiv \max\{X_1, \ldots, X_n\}$ . Recall that the distribution of the maximum  $M_n$  is given by  $H(x; n) = G(x)^n$ . Gabaix et al. (2016) show that if  $\zeta(G^{-1}(1-t)) \in RV_{\rho}^0$  for some  $\rho > -1$ , then in the limit as the number of draws n becomes large, we obtain

$$\mathbb{E}_{H}[\zeta(M_{n})] = \int_{\underline{x}}^{\overline{x}} \zeta(x) dH(x;n) \sim_{n \to \infty} \zeta\left(G^{-1}\left(1-\frac{1}{n}\right)\right) \Gamma(\rho+1)$$
(9)

where  $\Gamma : \mathbb{R} \to \mathbb{R}^+$  is the Gamma function defined by  $\Gamma(a) \equiv \int_0^\infty t^{a-1} e^{-t} dt$ .

As discussed in Gabaix et al. (2016), the probability that a draw from G exceeds the maximum  $M_n$  is approximately 1/n. Therefore, we have  $\mathbb{E}_H[1 - G(M_n)] \approx 1/n$ . This implies that  $\mathbb{E}_H[G(M_n)] \approx 1 - 1/n$  and thus  $\mathbb{E}_H[M_n] \approx G^{-1}(1 - 1/n)$  subject to a correction factor. More generally, we have  $\mathbb{E}_H[\zeta(M_n)] \approx \zeta(G^{-1}(1 - 1/n))$  subject to a correction factor. Expression (9) tells us the correction factor is  $\Gamma(\rho + 1)$ . If  $\rho \in (0, 1)$ , there is a downwards correction, otherwise the correction is upwards.<sup>23</sup>

The value of  $\rho$  depends on both the application and the underlying distribution G. Lemma 7 in Appendix C provides some useful facts about regular variation that are helpful for determining the value of  $\rho$  in specific applications. For example, if we consider just the maximum then  $\zeta(x) = x$  and Lemma 7 tells us that  $\rho = -\gamma$  if  $\overline{x} = +\infty$ .

#### 5.2. Random Number of Draws

Now suppose the number of draws  $N(\theta)$  is a random variable with mean  $\theta$  and define  $M_{N(\theta)} \equiv \max\{X_1, \ldots, X_{N(\theta)}\}$ . If the search technology  $P_n$  is mixed Poisson, the distribution of the maximum  $H_P(x;\theta)$  is given by Lemma 3. In the limit as the average number of draws  $\theta$  becomes large, we obtain the following result.

THEOREM 3—Extreme Value Outcomes: Suppose  $P_n$  is a mixed Poisson search technology with mean  $\theta$  and type distribution F. Assume that  $\zeta$  satisfies Assumption 3 and  $\zeta(G^{-1}(1-t)) \in RV_{\rho}^{0}$  for some  $\rho > -1$ . If  $\mathbb{E}_F[X^s]$  is finite for all s in a neighborhood of  $-\rho$ , then

$$\mathbb{E}_{H_P}[\zeta(M_{N(\theta)})] = \int_{\underline{x}}^{\overline{x}} \zeta(x) dH_P(x;\theta) \sim_{\theta \to \infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right) \Gamma(\rho+1) \underbrace{\mathbb{E}_F[X^{-\rho}]}_{\text{effect of heterogeneity}}$$

where  $\Gamma : \mathbb{R} \to \mathbb{R}^+$  is the Gamma function defined by  $\Gamma(a) \equiv \int_0^\infty t^{a-1} e^{-t} dt$ .

PROOF: See Appendix C.1.

We can provide a simple heuristic proof of Theorem 3 by exploiting the mixed Poisson representation of the search technology  $P_n$  and an existing result in Mangin (2024) for the Poisson special case. The formal proof of Theorem 3 can be found in Appendix C.1.

Q.E.D.

<sup>&</sup>lt;sup>23</sup>Since we assume that  $\zeta(G^{-1}(1-t)) \in RV_{\rho}^{0}$ , we know from Definition 2 that  $\zeta(G^{-1}(1-t))$  behaves like  $t^{\rho}$  in the limit as  $t \to 0$  where t = 1/n. Informally, the correction factor is upwards (downwards) if  $t^{\rho}$  is convex (concave) by Jensen's inequality. If  $\rho = 0$ , there is no correction.

#### 5.3. Heuristic Proof

Define  $H_{\tau}(x;\theta) \equiv H_P(x;\theta|\tau)$ , the distribution of the maximum *conditional on type*  $\tau$ . The distribution  $P_n$  conditional on  $\tau$  is Poisson with mean  $\theta\tau$ , so we can apply a result regarding extreme value outcomes for the Poisson special case in Mangin (2024), which says

$$\mathbb{E}_{H_{\tau}}[\zeta(M_{N(\theta)})] \sim_{\theta \to \infty} \zeta\left(G^{-1}\left(1 - \frac{1}{\theta\tau}\right)\right) \Gamma(\rho + 1).$$

Our assumption that  $\zeta(G^{-1}(1-t))$  is regularly varying at zero with index  $\rho$  implies

$$\mathbb{E}_{H_{\tau}}[\zeta(M_{N(\theta)})] \sim_{\theta \to \infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right) \Gamma(\rho+1)\tau^{-\rho}.$$

Next, using the fact that  $H_P(x;\theta) = \int_0^\infty H_\tau(x;\theta) dF(\tau)$ , we can write

$$\mathbb{E}_{H_P}[\zeta(M_{N(\theta)})] = \int_0^\infty \mathbb{E}_{H_\tau}[\zeta(M_{N(\theta)})]dF(\tau).$$

By properties of regular variation, we obtain the desired result:

$$\mathbb{E}_{H_P}[\zeta(M_{N(\theta)})] \sim_{\theta \to \infty} \zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right) \Gamma(\rho + 1) \underbrace{\mathbb{E}_F[X^{-\rho}]}_{\text{effect of heterogeneity}}.$$

#### 5.4. Effect of Heterogeneity

The term  $\mathbb{E}_F[X^{-\rho}]$  featured in Theorem 3 represents the *effect of heterogeneity* on extreme outcomes. This term is an additional correction factor. By Jensen's inequality, if  $\rho > 0$  and  $\tau^{-\rho}$  is convex, there is an upwards correction (i.e.  $\mathbb{E}_F[X^{-\rho}] > 1$ ), and if  $\rho < 0$  and  $\tau^{-\rho}$  is concave (since  $\rho > -1$ ), there is a downwards correction (i.e.  $\mathbb{E}_F[X^{-\rho}] < 1$ ).<sup>24</sup>

For any given value of  $\rho$ , the magnitude of the correction term  $\mathbb{E}_F[X^{-\rho}]$  is driven by the degree of heterogeneity of the type distribution F. If  $\rho > 0$  and  $\tau^{-\rho}$  is convex, the term  $\mathbb{E}_F[X^{-\rho}]$ is increasing in a mean-preserving spread of the type distribution F. However, if  $\rho < 0$  and  $\tau^{-\rho}$  is concave, the term  $\mathbb{E}_F[X^{-\rho}]$  is decreasing in a mean-preserving spread of F. Therefore, extreme value outcomes may be either increasing (for  $\rho > 0$ ) or decreasing (for  $\rho < 0$ ) in a mean-preserving spread of the type distribution F.

COROLLARY 4—Mean-Preserving Spread: Suppose that  $P_n$  is a mixed Poisson search technology with type distribution F and the assumptions of Theorem 3 hold for some  $\rho > -1$ .

- (i) If  $\rho > 0$ , extreme outcomes are increasing in a mean-preserving spread of F.
- (ii) If  $\rho < 0$ , extreme outcomes are decreasing in a mean-preserving spread of F.

(iii) If  $\rho = 0$ , extreme outcomes are not affected by a mean-preserving spread of F.

<sup>&</sup>lt;sup>24</sup>The reason why this correction moves in the opposite direction to the first correction is the following. Since we assume  $\zeta(G^{-1}(1-t)) \in RV_{\rho}^{0}$ , we know  $\zeta(G^{-1}(1-t))$  behaves like  $t^{\rho}$  in the limit as  $t \to 0$  where t = 1/n as for the case where n is deterministic. So, the first correction is still needed. However, n is now random, so an additional correction is required. Because  $\zeta(G^{-1}(1-1/n))$  behaves like  $n^{-\rho}$  in the limit as  $n \to \infty$ , the additional correction factor is upwards (downwards) if  $n^{-\rho}$  is convex (concave) by Jensen's inequality. If  $\rho = 0$ , there is no correction.

#### 5.5. *Examples*

We now present some examples of the behavior of extreme outcomes for different search technologies, a specific function  $\zeta$ , and a general underlying distribution G.

EXAMPLE 5.1—Poisson: If  $\zeta(x) = x^{\alpha}$  and  $\overline{x} = \infty$ , then Lemma 7 in Appendix C says  $G^{-1}(1-t) \in RV_{-\gamma}^{0}$  where  $\gamma$  is the tail index of G. Therefore,  $\zeta(G^{-1}(1-t)) \in RV_{\rho}^{0}$  where  $\rho = -\gamma \alpha$ . If  $P_n$  is a Poisson search technology and  $\alpha < 1/\gamma$ , then  $\rho > -1$  and

$$\mathbb{E}_{H_P}[M_{N(\theta)}^{\alpha}] \sim \left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)^{\alpha} \Gamma(1-\gamma\alpha).$$

For example, consider a generalized mean with exponent  $\varepsilon - 1$  where  $\varepsilon > 1$ . We choose this example because, in a model such as Melitz (2003) with monopolistic competition and heterogeneous firms with productivity distribution  $H_P$  and constant elasticity of substitution  $\varepsilon$ , this would represent aggregate productivity, which determines all other aggregate outcomes. Letting  $\alpha = \varepsilon - 1$  and assuming that  $\varepsilon < 1 + 1/\gamma$ , we obtain

$$\left(\int_{\underline{x}}^{\overline{x}} x^{\varepsilon-1} dH_P(x;\theta)\right)^{\frac{1}{\varepsilon-1}} \sim G^{-1} \left(1 - \frac{1}{\theta}\right) \left(\Gamma(1 - \gamma(\varepsilon - 1))\right)^{\frac{1}{\varepsilon-1}}.$$

This expression depends only on the distribution G and its tail index  $\gamma$ , plus the parameter  $\varepsilon$  from the function  $\zeta$ . There is no asymptotic effect of the search technology.

EXAMPLE 5.2—Negative Binomial: If  $\zeta(x) = x^{\alpha}$  and  $\overline{x} = \infty$ , then  $\zeta(G^{-1}(1-t)) \in RV^{0}_{\rho}$ where  $\rho = -\gamma \alpha$ . If  $P_{n}$  is a negative binomial search technology and  $r \in \mathbb{N} \setminus \{0\}$ , then F is the gamma distribution and  $\mathbb{E}_{F}[X^{-\rho}] = \frac{r^{\rho}\Gamma(r-\rho)}{\Gamma(r)}$ . If  $\alpha < 1/\gamma$ , then  $\rho > -1$  and

$$\mathbb{E}_{H_P}[M_{N(\theta)}^{\alpha}] \sim \left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)^{\alpha} \Gamma(1-\gamma\alpha) \underbrace{\frac{r^{-\gamma\alpha}\Gamma(r+\gamma\alpha)}{\Gamma(r)}}_{\text{effect of heterogeneity}}.$$

Again, letting  $\alpha = \varepsilon - 1$  and assuming that  $1 < \varepsilon < 1 + 1/\gamma$ , we obtain

$$\left(\int_{\underline{x}}^{\overline{x}} x^{\varepsilon-1} dH_P(x;\theta)\right)^{\frac{1}{\varepsilon-1}} \sim G^{-1} \left(1 - \frac{1}{\theta}\right) \left(\Gamma(1 - \gamma(\varepsilon - 1))\right)^{\frac{1}{\varepsilon-1}} \underbrace{r^{-\gamma} \left(\frac{\Gamma(r + \gamma(\varepsilon - 1))}{\Gamma(r)}\right)^{\frac{1}{\varepsilon-1}}}_{\text{effect of heterogeneity}}$$

Clearly, the effect of heterogeneity depends on the search technology through the parameter r, in addition to the tail index  $\gamma$  and the parameter  $\varepsilon$ . For example, if G is fat-tailed and  $\gamma > 0$ , we have  $\rho = -\gamma(\varepsilon - 1) < 0$  and a mean-preserving spread in the type distribution F has a negative effect on this extreme outcome by Corollary 4.

Next, we consider some applications of our general results to aggregate productivity, markups, and social networks. Before presenting these, we make the following assumption.

ASSUMPTION 4: The distribution G has tail index  $-1 \le \gamma < 1$  and the search technology  $P_n$  satisfies  $\mathbb{E}_F[X^s]$  is finite for all s in a neighborhood of  $-\rho$ .

Assumption 4 applies to all three of our applications (for different values of  $\rho$ ). This assumption ensures that we can apply Theorem 3 in all of these applications.

#### 6. APPLICATION: AGGREGATE PRODUCTIVITY

### 6.1. Environment

Suppose there is a continuum of measure one of firms. Time is discrete and continues forever. Firm-level productivity is influenced by an R&D process which generates new ideas. In each period, firms draw a number of new ideas from a common underlying distribution of ideas G which has unbounded upper support,  $[x, \infty) \subseteq \mathbb{R}_+$ .

The expected number of new ideas at a firm in any given period  $t \in \{1, 2, ...\}$  depends on both the aggregate R&D intensity  $\bar{\theta} \in \mathbb{R}^+$ , which is common across firms, and the firm-specific *R&D intensity*  $\tau$ , which is constant over time but heterogeneous across firms.<sup>25</sup> The distribution of R&D intensity across firms is given by a cdf *F* with support  $\tau \in [0, \infty)$ . For any given  $\tau$ , the number of ideas a firm receives is a Poisson random variable with mean equal to  $\tau\bar{\theta}$ . We assume the average R&D intensity  $\tau$  is one, so the average number of new ideas is  $\bar{\theta}$ .

We assume that an individual firm's productivity in period T is equal to the best idea that firm has received in any period  $t \in \{1, 2, ..., T\}$ . Applying Lemma 10, we know that the distribution of the cumulative number of draws received by a firm at time T is mixed Poisson with mean  $\hat{\theta}_T = T\bar{\theta}$  and mixing distribution F.<sup>26</sup> Let  $P_n^T$  denote this mixed Poisson distribution, which is our search technology in this application.

Suppose the distribution of R&D intensity is a gamma distribution with support  $\tau \in [0, \infty)$ and cdf  $F(\tau) = \frac{\gamma(r,r\tau)}{\Gamma(r)}$ . This implies the search technology  $P_n^T$  is negative binomial with parameter r and mean  $\hat{\theta}_T = T\bar{\theta}$ . A measure of dispersion is  $cv_F^2 = 1/r$ , so we refer to 1/r simply as the degree of R&D heterogeneity.<sup>27</sup>

Aggregate productivity in period T is equal to the average firm-level productivity in period T. That is, it is equal to the expected value of a firm's best idea where the total number of ideas is given by the search technology  $P_n$  with mean  $\hat{\theta}_T$ . That is,

$$y_P(\hat{\theta}_T) = \int_{\underline{x}}^{\infty} x dH_P(x; \hat{\theta}_T),$$

where  $H_P(x;\hat{\theta}_T)$  is the distribution of the maximum of a firm's ideas up until T.

Now consider the limit as  $T \to \infty$ . We have  $\hat{\theta}_T = T\bar{\theta} \to \infty$  for any  $\bar{\theta} \in \mathbb{R}^+$ . Therefore, we can apply our asymptotic results for large  $\theta$  in the limit as  $T \to \infty$ .

## 6.2. Results

Proposition 1 describes the asymptotic behavior of aggregate productivity  $y_P(\theta)$  and the impact of R&D heterogeneity. This result follows from Theorem 3 with  $\zeta(x) = x$ , which implies that  $\rho = -\gamma$  because  $\bar{x} = \infty$ . The results in Proposition 1 hold generally, not just for the negative binomial family. See Appendix D.1 for the general result which holds for any mixed Poisson search technology.

**PROPOSITION** 1—Aggregate Productivity: If  $P_n$  is negative binomial with parameter  $r \in \mathbb{N} \setminus \{0\}$ , the distribution of ideas G has tail index  $\gamma \in [0, 1)$  and  $\bar{x} = \infty$ , and  $r > -\gamma$ , then

<sup>&</sup>lt;sup>25</sup>We can interpret  $\bar{\theta}$  as representing spillovers from technological advances (from either private or public R&D) that benefit all firms in the economy. We can interpret firm-specific R&D intensity  $\tau$  as reflecting a firm's R&D expenditure and its effectiveness in generating new ideas.

<sup>&</sup>lt;sup>26</sup>The statement of Lemma 10 and its proof can be found in Appendix F.

<sup>&</sup>lt;sup>27</sup>The negative binomial family of mixed Poisson distributions was first used in Hausman et al. (1984) to model heterogeneity in R&D effectiveness across firms.

(i) Aggregate productivity is given by

$$y_P(\theta) \sim_{\theta \to \infty} G^{-1}\left(1 - \frac{1}{\theta}\right) \Gamma(1 - \gamma) \underbrace{\left(\frac{r^{-\gamma} \Gamma(r + \gamma)}{\Gamma(r)}\right)}_{\text{effect of R&D heterogeneiny}}.$$

(ii) If  $\gamma \in (0,1)$ , aggregate productivity  $y_P(\theta)$  is decreasing in R&D heterogeneity. (iii) If  $\gamma = 0$ , aggregate productivity  $y_P(\theta)$  does not depend on R&D heterogeneity.

PROOF: See Appendix D.1.

Proposition 1 implies that aggregate productivity is maximized when the R&D heterogeneity 1/r goes to zero, i.e. the search technology is Poisson. Intuitively, this is because there are diminishing marginal returns to the number of ideas discovered by a single firm. A planner would ideally want to equalize the number of new ideas across firms in order to maximize aggregate productivity.

Importantly, heterogeneity in R&D intensity influences not only the *level* of aggregate productivity but also the cross-sectional *distribution* of productivity across firms.

To obtain our next result, we apply Theorem 1 where  $P_0(z) = (\frac{r}{r+z})^r$ .

COROLLARY 5—Productivity Distribution: If  $P_n$  is negative binomial with parameter  $r \in \mathbb{N} \setminus \{0\}$ , the distribution of ideas G satisfies the conditions of Theorem 1 and has tail index  $\gamma \in [0, 1)$ , and  $r > -\gamma$ , the cdf of the (normalized) cross-sectional productivity distribution is

$$H_{\gamma,P}(x) = \left(\frac{r}{r+v_{\gamma}(x)}\right)^{\frac{1}{2}}$$

where  $v_{\gamma}(x) \equiv (1 + \gamma x)^{-1/\gamma}$  if  $\gamma \neq 0$  and  $e^{-x}$  if  $\gamma = 0$ .

### 6.3. Example: Fat-Tailed Distribution of Ideas

Suppose the underlying distribution of ideas is Pareto,  $G(x) = 1 - x^{-1/\gamma}$  with tail index  $\gamma \in (0, 1)$ . If all firms have the same cumulative number of draws n, i.e. if the search technology  $P_n$  is degenerate, the existing result in Gabaix et al. (2016) implies

$$y(n) \sim_{n \to \infty} n^{\gamma} \Gamma(1-\gamma)$$

If the search technology  $P_n$  is negative binomial with parameter  $r > -\gamma$ , then

$$y_P(\theta) \sim_{\theta \to \infty} \theta^{\gamma} \Gamma(1-\gamma) \underbrace{\left(\frac{r^{-\gamma} \Gamma(r+\gamma)}{\Gamma(r)}\right)}_{\text{effect of R&D heterogeneity}}$$

and the cdf of the (normalized) cross-sectional productivity distribution is

$$H_{\gamma,P}(x) = \left(\frac{r}{r+x^{-1/\gamma}}\right)^r.$$

EXAMPLE 6.1—Poisson: If  $P_n$  is a Poisson search technology, letting  $r \to \infty$  yields

$$y_P(\theta) \sim_{\theta \to \infty} \theta^{\gamma} \Gamma(1-\gamma)$$

Q.E.D.



FIGURE 3.—Cross-Sectional Firm Productivity Distribution (G is Pareto and  $\gamma = 1/4$ ).

and the cdf of the (normalized) cross-sectional productivity distribution is Fréchet:

$$H_{\gamma,P}(x) = e^{-x^{-1/\gamma}}.$$

EXAMPLE 6.2—Geometric: If  $P_n$  is a geometric search technology, setting r = 1 delivers

$$y_P(\theta) \sim_{\theta \to \infty} \theta^{\gamma} \Gamma(1-\gamma) \Gamma(1+\gamma).$$

In this case, the difference in the search technology has no effect if the tail index is  $\gamma = 0$ , but it leads to lower aggregate productivity if  $\gamma > 0$  because  $\Gamma(1 + \gamma) < 1$ .

The cdf of the (normalized) cross-sectional productivity distribution is

$$H_{\gamma,P}(x) = \frac{1}{1 + x^{-1/\gamma}}.$$

While the distribution is not Fréchet, the Pareto tail index  $\gamma$  is still inherited by the crosssectional productivity distribution, consistent with Corollary 3.

From Proposition 1, we know that aggregate productivity  $y_P(\theta)$  is strictly decreasing in R&D heterogeneity 1/r because  $\gamma \in (0, 1)$ . Therefore, aggregate productivity  $y_P(\theta)$  is highest when  $r \to \infty$  (Poisson) and lowest when r = 1 (geometric). For example, suppose that  $\gamma = 1/4$ . The ratio of aggregate productivity for the geometric versus Poisson search technology is  $\Gamma(1 + \gamma)$ , which is approximately 0.906 if  $\gamma = 1/4$ . This means the decline in aggregate productivity due to R&D heterogeneity is significant: around 9.4%.

The search technology also has a significant effect on cross-sectional productivity dispersion. Consider the measure of cross-sectional productivity dispersion  $cv_{H,P}$ , defined as the coefficient of variation of the distribution  $H_{\gamma,P}$ . Assume  $\gamma = 1/4$ . If the search technology  $P_n$  is Poisson, then  $cv_{H,P} = 0.42$ . However, if  $P_n$  is a geometric search technology then  $cv_{H,P} = 0.52$ . Therefore, the increase in cross-sectional productivity dispersion due to R&D heterogeneity is also significant: around 23%.

Figure 3 illustrates this example by depicting the Fréchet productivity distribution that arises when the search technology is Poisson (or deterministic), as well as the cross-sectional productivity distribution that arises when the search technology is geometric. When the search technology changes from Poisson to geometric, average productivity decreases and cross-sectional productivity dispersion increases. This is consistent with the first-order stochastic dominance result in Corollary 1. Observe that, regardless of whether the search technology is geometric or Poisson, the Pareto tail index  $\gamma$  is preserved as Corollary 3 suggests.

## 7. APPLICATION: MARKUPS

### 7.1. Environment

Consider a discrete choice model with random utility shocks. There is a continuum of measure one of consumers, and a continuum of firms. Each firm sells a single indivisible good, and each consumer has unit demand. Each consumer searches for firms and finds a number of firms  $n \in \{0, 1, 2, ...\}$  from which they can purchase.

In this application, the search technology reflects the fact that consumers are heterogeneous with respect to their *search intensity*. Some consumers search more intensely and find more firms on average, while others search less intensely and find fewer firms on average. We assume the actual number of firms a consumer with search intensity  $\sigma$  finds is a random variable which is Poisson with mean  $\sigma\theta$ . The average number of firms that a consumer finds is  $\theta$  because we assume the average search intensity is one.

Suppose the search intensity  $\sigma$  of a consumer is given by a Gamma distribution with support  $\sigma \in [0, \infty)$  and cdf  $F(\sigma) = \frac{\gamma(r, r\sigma)}{\Gamma(r)}$ . The distribution of n across consumers is therefore given by a negative binomial search technology  $P_n$  with parameter r. We can interpret  $cv_F^2 = 1/r$  as representing the degree of *consumer heterogeneity*.

Each consumer draws a random utility shock  $x_i$  from a distribution G for each firm i they find. This shock  $x_i$  represents the consumer's valuation of firm i's good. We assume the distribution G has support  $[\underline{x}, \overline{x}] \subseteq \mathbb{R}_+$  where  $\overline{x} \in \mathbb{R} \cup \{+\infty\}$ .

Suppose that the *n* firms a consumer finds set prices simultaneously, *after* observing the utility shocks  $(x_1, \ldots, x_n)$  and the number of competitors. Let  $M_n$  denote the maximum of  $(x_1, \ldots, x_n)$  and let  $S_n$  denote the second highest utility shock. In equilibrium, each consumer purchases from the firm which gives it the highest utility. The expected markup is the expected price minus marginal cost. If there is only one firm competing for a consumer, the expected markup is  $\mu(1) = \mathbb{E}_G(x)$ . If there are two or more firms, the expected markup  $\mu(n)$  is

$$\mu(n) = \mathbb{E}[M_n - S_n].$$

This type of pricing is often called "personalized pricing" or asymmetric Bertrand competition. For example, see Rhodes and Zhou (2024). While we focus on this type of pricing, our general result in Theorem 3 can also deliver the asymptotic markups for the wider class of random utility models studied in Gabaix et al. (2016). This includes the related models of Sattinger (1984) and Hart (1985), as well as the uniform pricing model of Perloff and Salop (1985). For some applications, this model is perhaps more realistic because firms set the same price for all consumers without observing consumers' utility shocks. As shown in Gabaix et al. (2016), however, all of these models have a common underlying logic and exhibit markups that are asymptotically *proportional* to the "personalized pricing" markup we consider here. The average markup across consumers, conditional on  $n \ge 1$  firms, is defined by

$$\mu_P(\theta) \equiv \sum_{n=1}^{\infty} \tilde{P}_n(\theta) \mu(n) \tag{10}$$

where  $\tilde{P}_n(\theta) \equiv \Pr(N(\theta) = n | n \ge 1)$ . If we assume  $\underline{x} = 0$ , the average markup is simply<sup>28</sup>

$$\mu_P(\theta) = \int_{\underline{x}}^{\overline{x}} \left(\frac{1 - G(x)}{g(x)}\right) dH_{\tilde{P}}(x;\theta).$$
(11)

We are interested in the asymptotic average markup that arises in the competitive limit where the average number of firms each consumer finds becomes large.

### 7.2. Results

Proposition 2 follows from Theorem 3 with  $\zeta(x) = \frac{1-G(x)}{g(x)}$ , which implies that  $\rho = -\gamma$ .<sup>29</sup> We provide an expression for the asymptotic average markup and describe the effect of an increase in the degree of consumer heterogeneity. These results hold generally, not just for the negative binomial family. See Appendix D.2 for the general result which holds for any mixed Poisson search technology.

**PROPOSITION 2**—Average Markup: If  $P_n$  is negative binomial with parameter  $r \in \mathbb{N} \setminus \{0\}$ , the distribution of utility shocks G has tail index  $\gamma \in [-1, 1)$ , and  $r > -\gamma$ , then (i) The average markup is given by

$$\mu_P(\theta) \sim_{\theta \to \infty} \frac{\Gamma(1-\gamma)}{\theta g \left( G^{-1} \left( 1 - \frac{1}{\theta} \right) \right)} \underbrace{\left( \frac{r^{-\gamma} \Gamma(r+\gamma)}{\Gamma(r)} \right)}_{\text{effect of consumer heterogeneity}}.$$
(12)

(ii) If  $\gamma \in (0,1)$ , the markup  $\mu_P(\theta)$  is decreasing in consumer heterogeneity. (iii) If  $\gamma \in [-1,0)$ , the markup  $\mu_P(\theta)$  is increasing in consumer heterogeneity. (iv) If  $\gamma = 0$ , the markup  $\mu_P(\theta)$  does not depend on consumer heterogeneity.

PROOF: See Appendix D.2.

If  $P_n$  is a Poisson search technology, letting  $r \to \infty$  in expression (12) yields the same expression for the asymptotic markup found in Gabaix et al. (2016). In general, there is a clear effect of consumer heterogeneity on markups (unless  $\gamma = 0$ ).

Greater consumer heterogeneity can either increase or decrease the average markup depending on the tail index  $\gamma$  of the underlying distribution of utility shocks. As shown in Gabaix et al. (2016), the elasticity of the markup  $\mu(n)$  with respect to n is asymptotically equal to the tail index  $\gamma$ . The tail index therefore determines whether the markup  $\mu(n)$  is asymptotically concave or convex, which governs the effect of greater dispersion.

 $<sup>\</sup>overline{2^{8}\text{If }\underline{x} > 0}$ , there is an extra term which goes to zero asymptotically. See Lemma 1 in Mangin (2024) for a general expression. The derivation uses the well-known result that  $\mathbb{E}[M_n - S_n] = \int_{\underline{x}}^{\underline{x}} \left(\frac{1 - G(x)}{g(x)}\right) dH(x;n)$  for  $n \ge 2$ .

<sup>&</sup>lt;sup>29</sup>Strictly speaking, expression (11) uses the conditional distribution  $\tilde{P}_n^-$  rather than the mixed Poisson search technology  $P_n$ , but we can use  $P_n$  when we apply Theorem 3 to (11) because  $P(N(\theta) = 1) \to 0$  as  $\theta \to \infty$ .



FIGURE 4.—Effect of Consumer Heterogeneity on Markups (G is uniform and  $\gamma = -1$ ).

If  $\gamma \in [-1, 0)$ , the markup  $\mu(n)$  is asymptotically decreasing and convex. This is because distributions with  $\gamma < 0$  are bounded above and the markup  $\mu(n) = \mathbb{E}[M_n - S_n]$ , i.e. the expected gap between the highest and second-highest utility shock, is asymptotically decreasing with n and converges to zero as n becomes large. In our environment where n is a random variable, the average markup  $\mu_P(\theta) = \mathbb{E}_P[\mu(n)]$  given by (10) is *increasing* in consumer heterogeneity because of the convexity of the markup  $\mu(n)$ .

If  $\gamma \in (0, 1)$ , the markup  $\mu(n)$  is asymptotically increasing and concave. This is because distributions with  $\gamma > 0$  are not bounded above and the markup  $\mu(n)$  is asymptotically increasing with n, although at a decreasing rate, as n becomes large. In our environment where n is a random variable, the average markup  $\mu_P(\theta) = \mathbb{E}_P[\mu(n)]$  is *decreasing* in consumer heterogeneity because of the concavity of the markup  $\mu(n)$ .

#### 7.3. Example: Uniform Distribution of Utility Shocks

Suppose the distribution of utility shocks is uniform, G(x) = x on [0, 1], which has tail index  $\gamma = -1$ . If all consumers contact the same number of firms n, i.e. if the search technology  $P_n$  is degenerate, we know from Gabaix et al. (2016) that

$$\mu(n) \sim_{n \to \infty} \frac{1}{n}$$

If the search technology  $P_n$  is negative binomial with parameter  $r > -\gamma$ , then

$$\mu_P(\theta) \sim_{\theta \to \infty} \frac{1}{\theta} \underbrace{\left(\frac{1}{1-1/r}\right)}_{\text{effect of consumer heterogeneity}}$$
(13)

Figure 4 illustrates the effect of consumer heterogeneity on markups. For any value of r, the average markup goes to zero in the limit as the expected number of firms each consumer finds

becomes large. As Figure 4 shows, the average markup is increasing in the degree of consumer heterogeneity 1/r, consistent with Proposition 2 because  $\gamma = -1$  for the uniform distribution. As we discussed, the reason why greater heterogeneity in search intensity increases the average markup is because  $\mu(n) = 1/n$  is convex. As Figure 4 shows, the average markup  $\mu_P(\theta)$  is lowest when  $r \to \infty$  (Poisson) and highest as  $r \to 1$ .

Our asymptotic markup expression (13) holds in the competitive limit as  $\theta \to \infty$ . However, it is a very good *approximation* of the exact markup (11) for finite but sufficiently large  $\theta$ . We might wonder, how large is sufficiently large? To answer this question for this example, we can compare the approximation for finite  $\theta$  given by (13) and the exact markup given by (11).

For higher values of r or lower values of consumer heterogeneity 1/r, the approximation is better. For example, if r = 5, the approximation is good (< 1% error) only if the average number of firms  $\theta \ge 20$ . However, if r = 10, the approximation is good (< 1% error) if  $\theta \ge 10$ and very good (< 0.05% error) if  $\theta \ge 20$ . For the Poisson search technology as  $r \to \infty$ , the approximation is very good (< 0.05% error) if the average number of firms  $\theta \ge 10$ .

EXAMPLE 7.1—Poisson: If  $P_n$  is a Poisson search technology, letting  $r \to \infty$  yields

$$\mu_P(\theta) \sim_{\theta \to \infty} \frac{1}{\theta}.$$

EXAMPLE 7.2—Negative Binomial: If  $P_n$  is negative binomial with r = 2, we obtain<sup>30</sup>

$$\mu_P(\theta) \sim_{\theta \to \infty} \frac{2}{\theta}.$$

The average markup for this example is twice as high as for the Poisson example, i.e. the increase in the average markup due to greater consumer heterogeneity is 100%.

## 8. APPLICATION: PEER EFFECTS IN SOCIAL NETWORKS

### 8.1. Environment

Consider a random social network of students. Each student is connected to a number of other students (their "friends") who are chosen uniformly at random from the population of students. Student outcomes are influenced by the amount of effort a student exerts, which is determined by peer effects, i.e. the influence of their friends' study effort on their own study effort. There is a large number of students.

There are two periods. In period one, each student exerts a level of study effort drawn from an exogenous distribution G with unbounded upper support,  $[\underline{x}, \infty) \subseteq \mathbb{R}_+$ .

In period two, each student observes their friends' effort levels in period one and chooses to study with the same effort as their most studious friend.<sup>31</sup> At the end of period two, students receive an outcome  $\phi(x)$  that is strictly increasing in their study effort x in period two, i.e.  $\phi'(x) > 0$ . The average student outcome is the expected value of  $\phi(x)$  across all students.

In this application, the search technology  $P_n$  is the *degree distribution* of the social network, i.e. the distribution of the number of friends across students.<sup>32</sup>

<sup>&</sup>lt;sup>30</sup>Note that we cannot consider the geometric example (where r = 1) because we require  $r > -\gamma$ .

<sup>&</sup>lt;sup>31</sup>While peer effects are often based on the *average* effort among friends, it is also common to consider the maximum effort among friends, e.g. see Boucher et al. (2024) for a discussion and a generalization. Instead of the maximum-effort friend, we could also consider the second-highest effort or the minimum-effort friend and our general results could still be applied.

<sup>&</sup>lt;sup>32</sup>While we keep this example simple, see Newman et al. (2001) for a development of the theory of random graphs that generalizes beyond the Poisson to arbitrary degree distributions.

25

We assume that students are heterogeneous with respect to their *popularity*  $\pi$ , which is distributed according to a cdf F with mean equal to one. For a student with popularity  $\pi$ , the exact number of friends is given by a Poisson distribution with mean  $\pi\theta$  where  $\theta$  is common across students. The average number of friends per student is  $\theta$ .

We assume that F is a Pareto distribution with support  $\pi \in [\pi_0, \infty)$  and cdf  $F(\pi) = 1 - \left(\frac{\pi}{\pi_0}\right)^{-1/\lambda}$  where  $\lambda \in (0,1)$  and  $\mathbb{E}_F[X] = \pi_0/(1-\lambda)$ . To ensure that  $\mathbb{E}_F[X] = 1$ , we set  $\pi_0 = 1 - \lambda$ . This distribution is useful because it has fat tails, with a higher parameter  $\lambda$  implying fatter tails. The resulting power-law search technology (or degree distribution) is mixed Poisson, as discussed in Campbell et al. (2024).<sup>33</sup>

Suppose that  $\phi(x) = bx^{\beta}$  where  $b, \beta > 0$  and  $\int_{\underline{x}}^{\overline{x}} |\phi(x)g(x)|dx$  is finite. The parameter  $\beta$  represents the curvature of outcomes as a function of study effort. We may have either  $\beta < 1$  ( $\phi$  concave),  $\beta > 1$  (convex), or  $\beta = 1$  ( $\phi$  linear). We assume that the curvature of student outcomes with respect to study effort is not too high, i.e.  $\beta < 1/\gamma$  where  $\gamma \ge 0$  is the tail index of the distribution G, the initial distribution of study effort.

Given there is a large number of students, the average outcome  $\phi_P(\theta)$  is

$$\phi_P(\theta) = \int_{\underline{x}}^{\infty} \phi(x) dH_P(x;\theta).$$

Now consider the limit as the network becomes dense, i.e. as the average number of friends  $\theta$  becomes large. Our question is: What is the asymptotic behavior of the average study outcome  $\phi_P(\theta)$  when the network becomes dense, i.e. as  $\theta \to \infty$ ?

Since we assume the popularity distribution F is Pareto, the search technology  $P_n$  is powerlaw and we have the following expression for any  $k < 1/\lambda$ :

$$\mathbb{E}_F[X^k] = \frac{(1-\lambda)^k}{1-k\lambda}$$

By varying the parameter  $\lambda$ , we can isolate the effect of the search technology. Since the mean is equal to one because we set  $\pi_0 = 1 - \lambda$ , an increase in  $\lambda$  is a mean-preserving spread and we thus refer to  $\lambda$  as the *popularity heterogeneity*.

## 8.2. Results

Proposition 3 follows from Theorem 3 with  $\zeta(x) = \phi(x)$ , which implies  $\rho = -\gamma\beta$ . These results hold more generally than the power-law family. See Appendix D.3 for the general result which holds for any mixed Poisson search technology and therefore for any distribution of popularity, not just the Pareto.

**PROPOSITION 3**—Average Outcome: If the distribution of popularity F is Pareto with tail index  $\lambda \in (0,1)$  and the distribution of initial study effort G has tail index  $\gamma \in [0,1)$ , then (i) The average student outcome is given by

$$\phi_P(\theta) \sim_{\theta \to \infty} b \left( G^{-1} \left( 1 - \frac{1}{\theta} \right) \right)^{\beta} \Gamma(1 - \beta \gamma) \underbrace{\frac{(1 - \lambda)^{\beta \gamma}}{1 - \beta \gamma \lambda}}_{\text{effect of popularity heterogeneity}}.$$

<sup>&</sup>lt;sup>33</sup>See Campbell et al. (2024) for a discussion of power-law degree distributions in random networks.

(ii) If  $\gamma \in (0,1)$ , average outcome  $\phi_P(\theta)$  is decreasing in popularity heterogeneity. (iii) If  $\gamma = 0$ , average outcome  $\phi_P(\theta)$  does not depend on popularity heterogeneity.

PROOF: See Appendix D.3.

Q.E.D.

Clearly, the effect of the search technology on the average student outcome depends on the degree of popularity heterogeneity  $\lambda$ . Specifically, greater heterogeneity in student popularity decreases the average outcome across students. Intuitively, this is because there are diminishing marginal returns to the number of friends of an individual student: an additional friend is more valuable to a less popular student with few friends than to a highly popular student with lots of friends. A planner seeking to maximize the average student outcome would therefore prefer less dispersion in popularity across students.

#### 8.3. Example: Fat-Tailed Distribution of Effort

Suppose the initial distribution of study effort is  $G(x) = 1 - \left(\frac{x}{x_0}\right)^{-1/\gamma}$  with tail index  $\gamma \in (0,1)$ . We set  $x_0 = 1 - \gamma$  to ensure that an increase in  $\gamma$  is a mean-preserving spread of the distribution of study effort. We refer to  $\gamma$  as the *effort heterogeneity*.

If all students have the same number of friends n, i.e. the search technology  $P_n$  is degenerate, the result from Gabaix et al. (2016) delivers the following expression:

$$\phi(n) \sim_{n \to \infty} b n^{\beta \gamma} (1 - \gamma)^{\beta} \Gamma(1 - \beta \gamma)$$

If students differ in their popularity and  $\lambda > 0$ , the average study outcome is

$$\phi_P(\theta) \sim_{\theta \to \infty} b \theta^{\beta \gamma} (1-\gamma)^{\beta} \Gamma(1-\beta \gamma) \underbrace{\frac{(1-\lambda)^{\beta \gamma}}{1-\beta \gamma \lambda}}_{effect of popularity heterogeneity}.$$

EXAMPLE 8.1—Power-law and  $\lambda \to 0$ : Let  $\phi(x) = x$ . If the popularity distribution F is Pareto with  $\pi_0 = 1 - \lambda$ , then in the limit as the tail index  $\lambda \to 0$ , we have

$$\phi_P(\theta) \sim_{\theta \to \infty} \theta^{\gamma} \Gamma(2 - \gamma),$$

where we use the fact that  $(1 - \gamma)\Gamma(1 - \gamma) = \Gamma(2 - \gamma)$  for the Gamma function.

EXAMPLE 8.2—Power-law and  $\lambda > 0$ : Let  $\phi(x) = x$ . If the popularity distribution F is Pareto with  $\pi_0 = 1 - \lambda$  and popularity heterogeneity  $\lambda > 0$ , then

$$\phi_P(\theta) \sim_{\theta \to \infty} \theta^{\gamma} \Gamma(2-\gamma) \frac{(1-\lambda)^{\gamma}}{1-\gamma \lambda}.$$

The average student outcome is decreasing in the popularity heterogeneity  $\lambda$ , but it is increasing in the effort heterogeneity  $\gamma$  for sufficiently large  $\theta$ .

For example, if  $\gamma = 1/2$  and  $\lambda = 1/2$  in the second example, the average student outcome is 5.7% lower than in the first example. Because an increase in  $\lambda$  is a mean-preserving spread of the popularity distribution F, the only difference between these two examples is the degree of heterogeneity in student popularity.

Heterogeneity in popularity clearly matters for student outcomes, even in the limit where the network becomes dense and the average number of friends becomes large. All else equal, a policy which leads to less popular students having more friends and more popular students having fewer friends would increase the average outcome. However, a policy which increases the average number of friends but also increases the degree of popularity dispersion may either increase or decrease the average outcome.

For example, let  $\theta$  denote the average number of friends and let  $\lambda \to 0$  before the policy change (i.e. the first example). Now let  $\theta(1 + \alpha)$  for  $\alpha > 0$  be the average number of friends and let  $\lambda > 0$  be the popularity heterogeneity after the policy change. Taking the limit as  $\theta$  becomes large, the ratio of student outcomes is

$$\frac{\phi_P^2(\theta(1+\alpha))}{\phi_P^1(\theta)} = \frac{(1+\alpha)^{\gamma}(1-\lambda)^{\gamma}}{1-\gamma\lambda}.$$

Therefore, the average student outcome is higher after the policy change if and only if

$$1 + \alpha > \frac{(1 - \gamma \lambda)^{1/\gamma}}{1 - \lambda}.$$

For example, if  $\lambda = 1/2$  and the effort heterogeneity is  $\gamma = 1/2$ , the policy change improves the average student outcome if and only if the average number of friends increases by more than 12.5%. Otherwise, the disadvantages of greater dispersion in student popularity more than offset the benefits of a higher average number of friends.

### 9. CONCLUSION

This paper provides some general results that allow us to apply extreme value theory in economic environments where agents are heterogeneous. We show that extreme value outcomes, and the nature of the extreme value distribution itself, depend not only on the underlying distribution of shocks and its tail index, but also on the *search technology*, which reflects heterogeneity in the expected number of draws across different types of agents.

In general, unless the search technology is Poisson, the extreme value distribution does not take any of the three standard forms (Fréchet, Gumbel, Weibull) and extreme value outcomes are influenced by the search technology. Interestingly, we find that extreme value outcomes may be either increasing or decreasing in the degree of heterogeneity across agents.

We consider some applications of our general results. We find that heterogeneity in R&D intensity across firms can have a quantitatively significant effect on both aggregate productivity and the degree of cross-sectional productivity dispersion. We also find that heterogeneity in consumers' search intensity can significantly affect the average markup and the effect may be either positive or negative. Finally, we apply our results to a simple model of peer effects in social networks and show that greater heterogeneity in student popularity can have a negative effect on the average student outcome.

While our asymptotic results strictly apply only in the limit as the expected number of draws becomes infinite, the expressions we derive for extreme value outcomes can also be used as *approximations* when the expected number of draws is finite but large enough. For example, we have seen that if the average number of firms a consumer finds is sufficiently large, our markup expression can be a very good approximation. Similarly, if the average number of friends in a social network is finite but sufficiently large, then our expression for the average student outcome is potentially a very good approximation.

We hope the simplicity, tractability, and flexibility of our expressions will make them useful for future applications. In particular, our new class of extreme value distributions can be used directly as a reduced-form generalization of the three standard extreme value distributions which incorporates additional dispersion without needing to use our microfoundation for this class. The class of extreme value distributions for the negative binomial family is particularly tractable and may be useful for empirical work in the future.

### APPENDIX A: PROOFS FOR SECTION 2

We first summarize some useful properties of the function  $P_0$  that will be used throughout.

LEMMA 4: If  $P_n$  is mixed Poisson, the function  $P_0 : \mathbb{R}^+ \to [0,1]$  is given by

$$P_0(\theta) = \int_0^\infty e^{-\theta\tau} dF(\tau).$$

The function  $P_0$  is continuous and infinitely differentiable for any  $\theta > 0$ , with  $P'_0(\theta) < 0$  and  $P''_0(\theta) > 0$  for all  $\theta > 0$ . We have (i)  $P_0(0) = 1$ ; (ii)  $P'_0(0) = -1$ ; and (iii)  $\lim_{\theta \to \infty} P_0(\theta) = 0$ .

PROOF: These properties follow from Assumption 2. See Theorem 1.4 in Schilling et al. (2012) for a proof that  $P_0$  is infinitely differentiable for any  $\theta > 0$ . Q.E.D.

**PROOF OF LEMMA 1:** Starting with Assumption 2, we have

$$P_n(\theta) = \int_0^\infty \frac{(\theta\tau)^n e^{-\theta\tau}}{n!} dF(\tau)$$

for some mixing distribution with cdf F. So, we have

$$\sum_{n=0}^{\infty} P_n(\theta) y^n = \sum_{n=0}^{\infty} y^n \int_0^\infty \frac{(\theta\tau)^n e^{-\theta\tau}}{n!} dF(\tau) = \int_0^\infty \sum_{n=0}^\infty \frac{(\theta y\tau)^n}{n!} e^{-\theta\tau} dF(\tau).$$

Therefore, the fact that  $\sum_{n=0}^{\infty} \frac{(\theta y \tau)^n}{n!} = e^{\theta y \tau}$  implies that

$$\sum_{n=0}^{\infty} P_n(\theta) y^n = \int_0^{\infty} e^{-\theta(1-y)\tau} dF(\tau).$$

By Lemma 4, we have  $P_0(z) = \int_0^\infty e^{-z\tau} dF(\tau)$ , thus  $\sum_{n=0}^\infty P_n(\theta) y^n = P_0(\theta(1-y))$ . Q.E.D.

### APPENDIX B: PROOFS FOR SECTION 4

## B.1. Proof of Theorem 1

PROOF OF THEOREM 1: Theorem 1 follows directly from the Barndorff-Nielsen (1964) theorem. The form below is based on Theorem 6.2.1 in Galambos (1987).

THEOREM—(Barndorff-Nielsen, 1964): Let  $X_1, \ldots, X_n$  be i.i.d. random variables and define  $M_n \equiv \max\{X_1, \ldots, X_n\}$ . Let  $a_n$  and  $b_n$  be sequences of normalizing constants such that the distribution of  $Z_n \equiv a_n M_n + b_n$  converges as  $n \to \infty$  to  $H_{\gamma}(x)$ . If the random variable  $N(\theta)/\theta$  converges in probability to a random variable X, then

$$\lim_{\theta \to \infty} \Pr(a_{\theta} M_{N(\theta)} + b_{\theta} \le x) = \int_{0}^{\infty} H_{\gamma}(x)^{\tau} d\Pr(X \le \tau).$$

By this theorem, if  $P_n$  is mixed Poisson with mixing distribution F, Lemma 5 implies

$$H_{\gamma,P}(x) = \lim_{\theta \to \infty} \Pr(a_{\theta} M_{N(\theta)} + b_{\theta} \le x) = \int_0^\infty H_{\gamma}(x)^{\tau} dF(\tau).$$

Given that  $H_{\gamma}(x) = e^{-v_{\gamma}(x)}$  by assumption of Theorem 1, this can be written as

$$H_{\gamma,P}(x) = \int_0^\infty e^{-v_\gamma(x)\tau} dF(\tau) = P_0(v_\gamma(x)).$$

This completes the proof of Theorem 1.

Lemma 5 is a known result, but we present a simple proof here for completeness.<sup>34</sup>

LEMMA 5: Suppose that  $N(\theta)$  is a random variable with mixed Poisson distribution  $P_n$ with mean  $\theta$  and mixing distribution F. In the limit as  $\theta \to \infty$ , the random variable  $N(\theta)/\theta$ converges in probability to the random variable X with distribution F.

PROOF: Suppose that  $N^{\tau}(\theta)$  is a Poisson distributed random variable with mean  $\theta \tau$  for some  $\tau > 0$ . Now consider the random variable  $Z = N^{\tau}(\theta)/\theta$ , where  $\mathbb{E}[Z] = \tau$  and  $Var(Z) = \theta \tau/\theta^2 = \tau/\theta$ . By Chebyshev's inequality, for any  $\epsilon > 0$  we have

$$\Pr(|Z - \mathbb{E}[Z]| \ge \epsilon) = \Pr(|N^{\tau}(\theta)/\theta - \tau| \ge \epsilon) \le \frac{Var(Z)}{\epsilon^2} = \frac{\tau}{\theta\epsilon^2}.$$

If  $N(\theta)$  is a mixed Poisson random variable with mean  $\theta$  and X denotes the random variable with mixing distribution F, then

$$\begin{split} \lim_{\theta \to \infty} \Pr(|N(\theta)/\theta - X| \ge \epsilon) &= \lim_{\theta \to \infty} \int_0^\infty \Pr(|N^\tau(\theta)/\theta - \tau| \ge \epsilon) dF(\tau) \\ &\leq \lim_{\theta \to \infty} \frac{1}{\theta \epsilon^2} \int_0^\infty \tau dF(\tau) = 0. \end{split}$$

Thus  $\lim_{\theta\to\infty} \Pr(|N(\theta)/\theta - X| \ge \epsilon) = 0$ , so  $N(\theta)/\theta$  converges in probability to X. Q.E.D.

## B.2. Proof of Theorem 2

PROOF OF THEOREM 2: Suppose the underlying distribution G is an extreme value distribution with cdf  $G(x) = P_0(v_{\gamma}(x))$  for some mixed Poisson search technology  $P_n$  and  $\gamma \in \mathbb{R}$ . Let  $X_1, \ldots, X_n$  be i.i.d. random variables with distribution G given by the cdf  $G(x) = P_0(v_{\gamma}(x))$ . Define the random variable  $M_n \equiv \max\{X_1, \ldots, X_n\}$ . Recall that  $H_{\gamma}(x) = e^{-v_{\gamma}(x)}$  where  $v_{\gamma}(x) \equiv (1 + \gamma x)^{-1/\gamma}$  if  $\gamma \neq 0$  and  $v_{\gamma}(x) \equiv e^{-x}$  if  $\gamma = 0$ .

The following useful result is Theorem 1.2.1 from de Haan and Ferreira (2006).

Q.E.D.

<sup>&</sup>lt;sup>34</sup>For closely related results in the literature, see Adell and de la Cal (1993) and Kuba and Panholzer (2016).

THEOREM—(de Haan and Ferreira, 2006): A distribution G is in the domain of attraction of the distribution  $H_{\gamma}$  if

(i) For  $\gamma > 0$ , we have

$$\lim_{t \to \bar{x}} \frac{1 - G(tx)}{1 - G(t)} = x^{-1/\gamma} \text{ for all } x > 0.$$

(*ii*) For  $\gamma < 0$ , we have

$$\lim_{t \to 0} \frac{1 - G(\bar{x} - tx)}{1 - G(\bar{x} - t)} = x^{-1/\gamma} \text{ for all } x > 0.$$

(iii) For  $\gamma = 0$ , there exists some positive function f > 0 such that

$$\lim_{t \to \bar{x}} \frac{1 - G(t + xf(t))}{1 - G(t)} = e^{-x} \text{ for all } x \in \mathbb{R}.$$

We can now prove Theorem 2 as follows. If  $\gamma > 0$ , then  $\bar{x} = +\infty$  and we have

$$\lim_{t \to \infty} \frac{1 - G(tx)}{1 - G(t)} = \lim_{t \to \infty} \frac{1 - P_0((1 + \gamma tx)^{-1/\gamma})}{1 - P_0((1 + \gamma t)^{-1/\gamma})}.$$

Applying L'Hôpital's rule, and using Lemma 4, it is straightforward to verify that

$$\lim_{t \to \infty} \frac{1 - P_0((1 + \gamma t x)^{-1/\gamma})}{1 - P_0((1 + \gamma t)^{-1/\gamma})} = \lim_{t \to \infty} x \left(\frac{1 + \gamma t x}{1 + \gamma t}\right)^{-1/\gamma - 1} = x^{-1/\gamma}.$$

If  $\gamma < 0$ , then  $\bar{x}$  is finite and we have

$$\lim_{t \to 0} \frac{1 - G(\bar{x} - tx)}{1 - G(\bar{x} - t)} = \lim_{t \to 0} \frac{1 - P_0((1 + \gamma(\bar{x} - tx))^{-1/\gamma})}{1 - P_0((1 + \gamma(\bar{x} - t))^{-1/\gamma})}.$$

Applying L'Hôpital's rule, and using the fact  $\lim_{x\to \bar{x}} G(x) = \lim_{x\to \bar{x}} P_0(v_\gamma(x)) = 1$ ,

$$\lim_{t \to 0} \frac{1 - P_0((1 + \gamma(\bar{x} - tx))^{-1/\gamma})}{1 - P_0((1 + \gamma(\bar{x} - t))^{-1/\gamma})} = \lim_{t \to 0} x \left(\frac{1 + \gamma(\bar{x} - tx)}{1 + \gamma(\bar{x} - t)}\right)^{-1/\gamma - 1} = x^{-1/\gamma}.$$

If  $\gamma = 0$ , then  $\bar{x} = +\infty$  and we have

$$\lim_{t \to \infty} \frac{1 - G(t + xf(t))}{1 - G(t)} = \lim_{t \to \infty} \frac{1 - P_0(v_\gamma(t + xf(t)))}{1 - P_0(v_\gamma(t))} = \lim_{t \to \infty} \frac{1 - P_0(e^{-(t + xf(t))})}{1 - P_0(e^{-t})}.$$
 (14)

Define the required function f by

$$f(t) \equiv \frac{\int_t^{\bar{x}} (1 - G(s)) ds}{1 - G(t)}$$

Applying L'Hôpital's rule to (14), and using Lemma 4 plus the fact that

$$\lim_{z \to 0} \frac{-P_0'(z)z}{1 - P_0(z)} = 1$$

by L'Hôpital's rule, we obtain  $\lim_{t\to \bar{x}} f(t) = 1$ ,  $\lim_{t\to \bar{x}} f'(t) = 0$ , and the following:

$$\lim_{t \to \infty} \frac{1 - P_0(e^{-(t+xf(t))})}{1 - P_0(e^{-t})} = \lim_{t \to \infty} e^{-xf(t)}(1 + xf'(t)) = e^{-x}.$$

Therefore, for any  $\gamma \in \mathbb{R}$ , the distribution G is in the domain of attraction of the extreme value distribution  $H_{\gamma}(x) = e^{-v_{\gamma}(x)}$  when the number of draws n is fixed.

Now suppose that  $N(\theta)$  is a random variable with the same mixed Poisson distribution  $P_n$ and mean  $\theta$ . Define  $M_{N(\theta)} \equiv \max\{X_1, \ldots, X_{N(\theta)}\}$ . By Theorem 1, there exist normalizing constants  $a_{\theta}$ ,  $b_{\theta}$  such that the sequence of normalized random variables  $Z_{N(\theta)} = a_{\theta}M_{N(\theta)} + b_{\theta}$ converges in distribution as  $\theta \to \infty$  to  $H_{\gamma,P}(x) = P_0(v_{\gamma}(x))$ . Therefore, the distribution G lies in its own domain of attraction, i.e. we have  $H_{\gamma,P}(x) = P_0(v_{\gamma}(x)) = G(x)$ . Q.E.D.

### B.3. Proof of Corollary 3

PROOF OF COROLLARY 3: Suppose that  $H_{\gamma,P}(x) = P_0(v_{\gamma}(x))$  for some mixed Poisson search technology  $P_n$  and some underlying distribution G with tail index  $\gamma \in \mathbb{R}$ . If  $H''_{\gamma,P}(x)$  exists and  $H'_{\gamma,P}(x) > 0$ , then we can apply Theorem 1.1.8 in de Haan and Ferreira (2006).

Suppose that  $H_{\gamma,P}$  has tail index equal to  $\gamma_H$ . By Definition 1, this implies that  $H''_{\gamma,P}(x)$  must exist.<sup>35</sup> To verify that  $H'_{\gamma,P}(x) > 0$ , differentiating  $H_{\gamma,P}(x) = P_0(v_{\gamma}(x))$  yields

$$H'_{\gamma,P}(x) = P'_0(v_{\gamma}(x))v'_{\gamma}(x).$$

We have  $P'_0(z) < 0$  by Lemma 4 and  $v'_{\gamma}(x) < 0$ , so  $H'_{\gamma,P}(x) > 0$ . We can now apply Theorem 1.1.8 in de Haan and Ferreira (2006), which says  $H_{\gamma,P}$  is in the domain of attraction of  $H_{\gamma_H}$  where  $H_{\gamma_H}(x) = e^{-v_{\gamma_H}(x)}$  if the number of draws is fixed. Therefore, our Theorem 1 tells us that  $H_{\gamma,P}$  is in the domain of attraction of  $H_{\gamma_H,P}$  if the number of draws is random and given by the search technology  $P_n$ . Next, our Theorem 2 says that  $H_{\gamma,P}$  lies is in its own domain of attraction when the number of draws is random and given by the search technology  $P_n$ .

Therefore,  $H_{\gamma,P}$  is in the domain of attraction of both  $H_{\gamma_H,P}$  and  $H_{\gamma,P}$ . As discussed in de Haan and Ferreira (2006), domains of attraction are unique, which implies that  $\gamma_H = \gamma$ . Q.E.D.

### APPENDIX C: PROOFS FOR SECTION 5

#### C.1. Proof of Theorem 3

We provide two alternative proofs of Theorem 3. The first proof is simpler and formalizes the heuristic proof in the main text using the existing Poisson result from Mangin (2024). This proof does not require the assumption regarding boundedness of  $\zeta$  over any closed interval stated in Assumption 3.

The second proof is less intuitive but establishes the result in Theorem 3 without using the Poisson result from Mangin (2024), although it does require the condition in Assumption 3 regarding boundedness of  $\zeta$  over any closed interval.

<sup>&</sup>lt;sup>35</sup>In order for  $H_{\gamma,P}$  to have tail index  $\gamma_H \in \mathbb{R}$  by Definition 1, it is necessary (but not sufficient) that  $H_{\gamma,P}$  is twice-differentiable. This will be true if  $P_0$  is twice-differentiable, which is true for any  $\theta > 0$  by Lemma 4.

SIMPLE PROOF OF THEOREM 3: Theorem 3 can be expressed in the following way:

$$\lim_{\theta \to \infty} \frac{\mathbb{E}_{H_P}[\zeta(x)]}{\zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right)} = \Gamma(\rho + 1)\mathbb{E}_F[X^{-\rho}].$$

Using the fact that  $\mathbb{E}_{H_P}[\zeta(x)] = \int_0^\infty \mathbb{E}_{H_\tau}[\zeta(x)] dF(\tau)$ , this is equivalent to

$$\lim_{\theta \to \infty} \int_0^\infty \frac{\mathbb{E}_{H_\tau}[\zeta(x)]}{\zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right)} dF(\tau) = \Gamma(\rho + 1)\mathbb{E}_F[X^{-\rho}]$$

where  $H_{\tau}(x;\theta) \equiv H_P(x;\theta|\tau)$ . Rearranging the left-hand side, we can write

$$\lim_{\theta \to \infty} \int_0^\infty \frac{\mathbb{E}_{H_\tau}[\zeta(x)]}{\zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right)} dF(\tau) = \lim_{\theta \to \infty} \frac{\mathbb{E}_{H_1}[\zeta(x)]}{\zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right)} \int_0^\infty \frac{\mathbb{E}_{H_\tau}[\zeta(x)]}{\mathbb{E}_{H_1}[\zeta(x)]} dF(\tau).$$

By the mixed Poisson assumption, the distribution  $P_n$  conditional on type  $\tau$  is Poisson with mean  $\theta\tau$ . Therefore, we can use Theorem B1 in Mangin (2024) for the Poisson special case, which implies the following:

$$\mathbb{E}_{H_{\tau}}[\zeta(x)] \sim_{\theta \to \infty} \zeta\left(G^{-1}\left(1 - \frac{1}{\theta\tau}\right)\right) \Gamma(\rho + 1).$$
(15)

Applying the above result (15) for  $\tau = 1$ , we get

$$\lim_{\theta \to \infty} \int_0^\infty \frac{\mathbb{E}_{H_\tau}[\zeta(x)]}{\zeta \left( G^{-1} \left( 1 - \frac{1}{\theta} \right) \right)} dF(\tau) = \Gamma(\rho + 1) \lim_{\theta \to \infty} \int_0^\infty \frac{\mathbb{E}_{H_\tau}[\zeta(x)]}{\mathbb{E}_{H_1}[\zeta(x)]} dF(\tau).$$

By (3), we have  $H_{\tau}(x;\theta) = e^{-\theta \tau (1-G(x))}$ . Defining  $k(\theta \tau) \equiv \mathbb{E}_{H_{\tau}}[\zeta(x)]$ , we have

$$\lim_{\theta \to \infty} \frac{k(\theta\tau)}{k(\theta)} = \lim_{\theta \to \infty} \frac{\mathbb{E}_{H_{\tau}}[\zeta(x)]}{\mathbb{E}_{H_{1}}[\zeta(x)]} = \lim_{\theta \to \infty} \frac{\zeta\left(G^{-1}\left(1 - \frac{1}{\theta\tau}\right)\right)}{\zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right)} = \tau^{-\rho}$$

where  $k(t) \in RV_{-\rho}^{\infty}$ . Defining  $\phi_{\theta}(\tau) \equiv \frac{k(\theta \tau)}{k(\theta)}$ , our result follows immediately from Lemma 6 because Lemma 6 implies that

$$\lim_{\theta \to \infty} \int_0^\infty \frac{\mathbb{E}_{H_\tau}[\zeta(x)]}{\mathbb{E}_{H_1}[\zeta(x)]} dF(\tau) = \int_0^\infty \lim_{\theta \to \infty} \frac{\mathbb{E}_{H_\tau}[\zeta(x)]}{\mathbb{E}_{H_1}[\zeta(x)]} dF(\tau) = \mathbb{E}_F[X^{-\rho}].$$
*Q.E.D.*

LEMMA 6: If the distribution of types F has minimum type  $\underline{\tau} \ge 0$ , then

$$\lim_{\theta \to \infty} \int_0^\infty \phi_\theta(\tau) dF(\tau) = \int_0^\infty \lim_{\theta \to \infty} \phi_\theta(\tau) dF(\tau).$$
(16)

PROOF: By the dominated convergence theorem, it suffices to prove there exists a function  $g(\tau)$  such that  $|\phi_{\theta}(\tau)| \leq g(\tau)$  for *F*-almost all  $\tau$  and  $\int_{0}^{\infty} g(\tau)dF(\tau) < \infty$ . Given that  $k(t) \in RV_{-\rho}^{\infty}$ , we can apply Potter's theorem (Theorem 1.5.6 in Bingham et al. (1987)), which says that, for any A > 1 and  $\varepsilon > 0$ , there exists  $\theta(A, \varepsilon)$  such that

$$\frac{k(y)}{k(x)} \le A \max\left\{ \left(\frac{y}{x}\right)^{-\rho+\varepsilon}, \left(\frac{y}{x}\right)^{-\rho-\varepsilon} \right\}$$

for any  $x, y \ge \theta(A, \varepsilon)$ . Consider any  $\tau > 0$ . Letting  $y = \theta \tau$  and  $x = \theta$ , we have  $\phi_{\theta}(\tau) \le A \max\{\tau^{-\rho+\varepsilon}, \tau^{-\rho-\varepsilon}\}$  for  $x, y \ge \theta(A, \varepsilon)$ . Given  $\varepsilon > 0$ , define  $g(\tau) = A(\tau^{-\rho+\varepsilon} + \tau^{-\rho-\varepsilon})$ .

By Potter's theorem, we have  $\phi_{\theta}(\tau) \leq g(\tau)$  for  $\theta$  sufficiently large that  $\theta \geq \theta(A, \varepsilon)$  and  $\theta \tau \geq \theta(A, \varepsilon)$ . Since k is positive, we have  $|\phi_{\theta}(\tau)| \leq g(\tau)$  for any  $\tau > 0$ . If  $\underline{\tau} > 0$ , this is clearly sufficient. If  $\underline{\tau} = 0$ , we have  $|\phi_{\theta}(\tau)| \leq g(\tau)$  for F-almost all  $\tau$  because F is continuous and there is no mass point at zero. Given that we assume that  $\mathbb{E}_F[X^s]$  is finite for all s in a neighborhood of  $-\rho$ , we have  $\int_0^{\infty} g(\tau) dF(\tau) < \infty$  and (16) is proven. Q.E.D.

PROOF OF THEOREM 3: Defining  $k : \mathbb{R}^+ \to [0,1]$  by  $k(z) = -P_0'(z)$ , we have

$$h_P(x;\theta) = g(x)\theta k(\theta(1-G(x))).$$

Therefore, we have

$$\int_{\underline{x}}^{\overline{x}} \zeta(x) h_P(x;\theta) dx = \int_{\underline{x}}^{\infty} \zeta(x) g(x) \theta k(\theta(1-G(x))) dx.$$
(17)

For any given  $\theta$ , the integral on the right-hand side of (17) is finite because Assumption 3 says that  $\int_x^x |\zeta(x)g(x)| \, dx < \infty$  and  $k : \mathbb{R}^+ \to [0, 1]$ . Letting  $x = G^{-1} \left(1 - \frac{1}{t}\right)$ , we obtain

$$\int_{\underline{x}}^{\overline{x}} \zeta(x) h_P(x;\theta) dx = \int_1^\infty k\left(\frac{\theta}{t}\right) \frac{\theta}{t} \zeta\left(G^{-1}\left(1-\frac{1}{t}\right)\right) \frac{dt}{t}.$$

Define  $k_0 : \mathbb{R}^+ \to \mathbb{R}^+$  by  $k_0(z) \equiv zk(z)$  and define  $\widetilde{\zeta} : [0, \infty) \to \mathbb{R}$  by  $\widetilde{\zeta}(t) = \zeta \left( G^{-1} \left( 1 - \frac{1}{t} \right) \right)$  for  $t \in [1, \infty)$  and  $\widetilde{\zeta}(t) = 0$  for  $t \in [0, 1)$ . We can thus write:

$$\int_{\underline{x}}^{\overline{x}} \zeta(x) h_P(x;\theta) dx = \int_0^\infty k_0\left(\frac{\theta}{t}\right) \widetilde{\zeta}(t) \frac{dt}{t}$$

Rewriting the above, we have  $\int_{\underline{x}}^{\overline{x}} \zeta(x) h_P(x;\theta) dx = (k_0 \overset{M}{*} \widetilde{\zeta})(\theta)$  where  $(\phi \overset{M}{*} f)(\theta)$  denotes the Mellin convolution of  $\phi$  and f, evaluated at  $\theta$ , defined by

$$(\phi \stackrel{M}{*} f)(\theta) \equiv \int_{0}^{\infty} \phi\left(\frac{\theta}{t}\right) f(t) \frac{dt}{t}.$$

We now apply Theorem 4.1.6 from Bingham et al. (1987), which was originally proved in Arandelovic (1976). This result says that if (i)  $f : [0, \infty) \to \mathbb{R}$  is measurable, (ii)  $f(t) \in RV_{v}^{\infty}$ , (iii) there exists  $\sigma, \tau \in \mathbb{R}$  such that  $v \in (\sigma, \tau)$  and for all  $s \in [\sigma, \tau]$ , the Mellin transform of  $\phi$ (defined by  $\check{\phi}(-s) \equiv \int_{0}^{\infty} t^{-s-1} \phi(t) dt$ ) is finite, and (iv)  $f(t)/t^{\sigma}$  is bounded on (0, t] for any t > 0, then

$$(\phi^{M} * f)(\theta) \sim_{\theta \to \infty} \breve{\phi}(-\upsilon) f(\theta).$$

We verify that the conditions for applying this theorem hold when  $f = \tilde{\zeta}$ ,  $v = -\rho$ , and  $\phi = k_0$ . (i) By Assumption 3, we know  $\zeta$  is measurable, so  $\tilde{\zeta}$  is measurable. (ii) By assumption,  $\zeta(G^{-1}(1-t)) \in RV^0_{\rho}$ , which implies  $\widetilde{\zeta}(t) \in RV^{\infty}_{-\rho}$  since  $\widetilde{\zeta}(t) = \zeta(G^{-1}(1-1/t))$  for  $t \in C^{-1}(1-1/t)$  $[1,\infty)$ . (iii) As shown below, the Mellin transform  $\check{k}_0(-s)$  equals  $\Gamma(1-s)\mathbb{E}_F(X^s)$ . By our assumption that  $\mathbb{E}_F(X^s)$  is finite for all s in a neighbourhood of  $-\rho$ , there exists  $\sigma, \tau \in \mathbb{R}$  such that  $-\rho \in (\sigma, \tau)$  and for all  $s \in [\sigma, \tau]$ ,  $\mathbb{E}_F(X^s)$  is finite. Moreover, since  $\rho > -1$  by assumption,  $-\rho < 1$  and therefore we can choose  $\tau < 1$  so that  $\Gamma(1-s)$  is finite for all  $s \in [\sigma, \tau]$  (since the Gamma function is finite over  $(0,\infty)$ ). Thus,  $k_0(-s) = \Gamma(1-s)\mathbb{E}_F(X^s)$  is finite for all  $s \in [\sigma, \tau]$ . (iv) Since  $\widetilde{\zeta}(t) = 0$  for all t < 1, (iv) is equivalent to boundedness of  $\widetilde{\zeta}$  on any interval [1,t] for t > 1 (because  $t^{-\sigma}$  is bounded on any such interval). By definition of  $\tilde{\zeta}$ , this is equivalent to boundedness of  $\zeta$  on every closed interval  $[\underline{x}, x]$ , which is assumed to be true by Assumption 3.

Applying Theorem 4.1.6 from Bingham et al. (1987), we obtain

$$(k_0 \stackrel{M}{*} \widetilde{\zeta})(\theta) \sim_{\theta \to \infty} \breve{k}_0(\rho) \widetilde{\zeta}(\theta).$$

Therefore, given that  $\breve{k}_0(\rho) = \int_0^\infty t^\rho k(t) dt$  and  $\widetilde{\zeta}(\theta) = \zeta \left( G^{-1} \left( 1 - \frac{1}{\theta} \right) \right)$ , we have

$$\int_{\underline{x}}^{\overline{x}} \zeta(x) h_P(x;\theta) dx \sim_{\theta \to \infty} \zeta \left( G^{-1} \left( 1 - \frac{1}{\theta} \right) \right) \int_0^\infty t^\rho k(t) dt.$$

Since  $k(t) = -P'_0(t)$ , we have  $\int_0^\infty t^\rho k(t)dt = -\int_0^\infty t^\rho P'_0(t)dt$ . Using Lemma 4, we obtain the following:

$$-\int_0^\infty t^{a-1} P_0'(t) dt = \int_0^\infty t^{a-1} \int_0^\infty u e^{-tu} dF(u) dt.$$

Finally, we perform the change of variables v = tu to get

$$\begin{split} \int_0^\infty t^{a-1} \int_0^\infty u e^{-tu} dF(u) dt &= \int_0^\infty (v/u)^{a-1} \int_0^\infty e^{-v} dF(u) dv \\ &= \int_0^\infty v^{a-1} e^{-v} dv \int_0^\infty u^{1-a} dF(u) \end{split}$$

which equals  $\Gamma(a)\mathbb{E}_F(X^{1-a})$  by definition. Setting  $a = \rho + 1$ , Theorem 3 is proven. O.E.D.

The following lemma is useful for applying Theorem 3. See Lemma 1 and Lemma A1 in Gabaix et al. (2016) and Proposition 1.5.7 in Bingham et al. (1987).

LEMMA 7—Regular Variation: If the distribution G is well-behaved with tail index  $\gamma < 1$ ,

(i) We have  $g(G^{-1}(1-t)) \in RV_{\rho}^{0}$  where  $\rho = \gamma + 1$ .

- (i) We have  $g(G (1-t)) \in RV_{\rho}^{\circ}$  where  $\rho = \gamma + 1$ . (ii) If  $\overline{x} = \infty$ , then  $G^{-1}(1-t) \in RV_{\rho}^{\circ}$  where  $\rho = -\gamma$ . (iii) If  $\overline{x} < \infty$ ,  $\overline{x} G^{-1}(1-t) \in RV_{\rho}^{\circ}$  where  $\rho = -\gamma$ . (iv) If  $k(t) \in RV_{\rho_{1}}^{\circ}$  then  $k(t)^{\alpha} \in RV_{\rho}^{\circ}$  where  $\rho = \alpha\rho_{1}$  for any  $\alpha \in \mathbb{R}$ . (v) If  $k_{1}(t) \in RV_{\rho_{1}}^{\circ}$  and  $k_{2}(t) \in RV_{\rho_{2}}^{\circ}$  then  $k_{1}(t)k_{2}(t) \in RV_{\rho}^{\circ}$  where  $\rho = \rho_{1} + \rho_{2}$ . (vi) If  $k_{1}(t) \in RV_{\rho_{1}}^{\circ}$  and  $k_{2}(t) \in RV_{\rho_{2}}^{\circ}$  then  $k_{1}(t) + k_{2}(t) \in RV_{\rho}^{\circ}$  where  $\rho = \max\{\rho_{1}, \rho_{2}\}$ .

(vii) We have  $k(t) \in RV_{\rho}^{0}$  if and only if  $\hat{k}(t) \equiv k(1/t) \in RV_{-\rho}^{\infty}$ .

#### **APPENDIX D: PROOFS FOR APPLICATIONS**

## D.1. Proof of Proposition 1

PROOF OF PROPOSITION 1: This result follows directly from Lemma 8 together with the more general Proposition 4, which holds for any mixed Poisson search technology. *Q.E.D.* 

PROPOSITION 4—Aggregate Productivity: If  $P_n$  is mixed Poisson, the distribution of ideas G has tail index  $\gamma \in [0, 1)$  and  $\bar{x} = \infty$ , and Assumption 4 holds for  $\rho = -\gamma$ , then

(i) Aggregate productivity is given by

$$y_P(\theta) \sim_{\theta \to \infty} G^{-1}\left(1 - \frac{1}{\theta}\right) \Gamma(1 - \gamma) \underbrace{\mathbb{E}_F[X^{\gamma}]}_{\text{effect of R&D betermagnetic}}$$

(ii) If  $\gamma \in (0,1)$ , then  $y_P(\theta)$  is decreasing in a mean-preserving spread of F. (iii) If  $\gamma = 0$ , then  $y_P(\theta)$  is not affected by a mean-preserving spread of F.

PROOF: It is clear that  $\zeta$  satisfies Assumption 3 because  $\int_{\underline{x}}^{\overline{x}} |xg(x)| dx$  is finite by Assumption 1. Moreover, since we assume  $\overline{x} = \infty$ , Lemma 7 implies that we have  $G^{-1}(1-t) \in RV_{\rho}^{0}$  where  $\rho = -\gamma$  and  $\gamma$  is the tail index of G. Finally,  $\rho > -1$  because  $\gamma < 1$ . Parts (ii) and (iii) follow from Corollary 4.

LEMMA 8: If  $P_n$  is negative binomial with parameter  $r \in \mathbb{N} \setminus \{0\}$  where  $r > -\gamma$ , then

$$\mathbb{E}_F[X^{\gamma}] = \frac{r^{-\gamma} \Gamma(r+\gamma)}{\Gamma(r)}.$$
(18)

(i) If γ ∈ (0,1), then E<sub>F</sub>[X<sup>γ</sup>] is decreasing in 1/r.
(ii) If γ ∈ [-1,0), then E<sub>F</sub>[X<sup>γ</sup>] is increasing in 1/r.

PROOF: The mixing distribution F is a gamma distribution given by  $F(x) = \frac{\gamma(r,rx)}{\Gamma(r)}$  and it is known that  $\mathbb{E}_F[X^{-\rho}] = \frac{r^{\rho}\Gamma(r-\rho)}{\Gamma(r)}$ . Expression (18) follows by setting  $\rho = -\gamma$ . To prove parts (i) and (ii), we can apply Corollary 4. We need only show that an increase in 1/r or equivalently a decrease in r is a mean-preserving spread of the distribution F.

Let  $r \in \mathbb{N} \setminus \{0\}$  and let  $F_r(x)$  denote  $\frac{\gamma(r,rx)}{\Gamma(r)}$ . We need to show that  $F_r$  is a mean-preserving spread of  $F_{r+k}$  for any  $k \ge 1$ . Given that  $F_r$  and  $F_{r+k}$  have the same mean, it suffices to prove that  $F_r$  and  $F_{r+k}$  have a single crossing, i.e. that there exists  $x^* \in (0,\infty)$  such that  $F_{r+k}(x) \le F_r(x)$  for all  $x < x^*$  and  $F_{r+k}(x) \ge F_r(x)$  for all  $x > x^*$ .

For any given  $k \ge 1$ , we have  $F_{r+k}(x) \le F_r(x)$  if and only if

$$\phi_k(r,x) \equiv \gamma(r+k,(r+k)x) - \frac{\Gamma(r+k)}{\Gamma(r)}\gamma(r,rx) \le 0.$$

Differentiating  $\phi_k(r,x)$  using the fact that  $\frac{\partial}{\partial y}\gamma(s,y) = y^{s-1}e^{-y}$ , we obtain

$$\frac{d}{dx}\phi_k(r,x) = x^{r-1}e^{-rx}(r+k)^{r+k}\left(x^k e^{-kx} - \frac{\Gamma(r+k)}{\Gamma(r)}\frac{r^r}{(r+k)^{r+k}}\right).$$

Given any  $k \ge 1$ , it can be verified that  $\frac{d}{dx}\phi_k(r,x) = 0$  at exactly two interior points. Given that  $\phi_k(r,0) = 0$  and  $\frac{d}{dx}\phi_k(r,x) < 0$  for small x > 0, plus  $\lim_{x\to\infty} \phi_k(r,x) = 0$  and  $\lim_{x\to\infty} \frac{d}{dx} \phi_k(r,x) < 0, \text{ there exists } x^* \in (0,\infty) \text{ such that } \phi_k(r,x) < 0 \text{ for all } x < x^* \text{ and } \phi_k(r,x) > 0 \text{ for all } x > x^*, \text{ which gives us the desired result.}$ 

### D.2. Proof of Proposition 2

PROOF OF PROPOSITION 2: This result follows directly from Lemma 8 together with the more general Proposition 5, which holds for any mixed Poisson search technology. *Q.E.D.* 

PROPOSITION 5—Average Markup: If  $P_n$  is mixed Poisson, the distribution of utility shocks G has tail index  $\gamma \in [-1, 1)$ , and Assumption 4 holds for  $\rho = -\gamma$ , then

(i) The average markup is given by

$$\mu_P(\theta) \sim_{\theta \to \infty} \frac{\Gamma(1-\gamma)}{\theta g \left( G^{-1} \left( 1 - \frac{1}{\theta} \right) \right)} \underbrace{\mathbb{E}_F[X^{\gamma}]}_{\text{effect of consumer heterogeneity}}$$

(ii) If  $\gamma \in (0,1)$ , the markup  $\mu_P(\theta)$  is decreasing in a mean-preserving spread of F. (iii) If  $\gamma \in [-1,0)$ , the markup  $\mu_P(\theta)$  is increasing in a mean-preserving spread of F. (iv) If  $\gamma = 0$ , the markup  $\mu_P(\theta)$  is not affected by a mean-preserving spread of F.

 $\begin{array}{l} \text{PROOF: First, } \zeta = \frac{1-G}{g} \text{ satisfies Assumption 3. Moreover } \zeta(G^{-1}(1-t)) = \frac{1-G(G^{-1}(1-t))}{g(G^{-1}(1-t))} = \frac{t}{g(G^{-1}(1-t))} \in RV_{\rho}^{0} \text{ where } \rho = -\gamma \text{ because } t \in RV_{1}^{0} \text{ and } g(G^{-1}(1-t)) \in RV_{\gamma+1}^{0} \text{ by Lemma } 7. \text{ Finally, } \rho > -1 \text{ since } \gamma < 1. \text{ Parts (ii), (iii), (iv) follow from Corollary 4.} \end{array}$ 

#### D.3. Proof of Proposition 3

PROOF OF PROPOSITION 3: We can use Proposition 6 to prove Proposition 3. Proposition 6 holds for any mixed Poisson search technology, not just the power law family. Part (i) of Proposition 3 follows from Proposition 6 plus the fact that  $\mathbb{E}_F[X^k] = \frac{(1-\lambda)^k}{1-k\lambda}$ . For part (ii), an increase in  $\lambda$  is a mean-preserving spread of F, so Corollary 4 applies. Part (iii) is clear. Q.E.D.

**PROPOSITION 6**—Average Outcome: If  $P_n$  is mixed Poisson, the distribution of initial study effort G has tail index  $\gamma \in [0,1)$  and  $\bar{x} = \infty$ , and Assumption 4 holds for  $\rho = -\gamma\beta$ , then (i) The average student outcome is given by

$$\phi_P(\theta) \sim_{\theta \to \infty} b\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right)^{\beta} \Gamma(1 - \beta\gamma) \underbrace{\mathbb{E}_F[X^{\gamma\beta}]}_{effect of popularity heterogeneity}$$

(ii) If  $\gamma \in (0,1)$ , then  $\phi_P(\theta)$  is decreasing in a mean-preserving spread of F. (iii) If  $\gamma = 0$ , then  $\phi_P(\theta)$  is not affected by a mean-preserving spread of F.

PROOF: It is clear that  $\zeta = \phi$  satisfies Assumption 3 since we assume that  $\int_{\underline{x}}^{\overline{x}} |\phi(x)g(x)| dx$ is finite. Moreover, because  $\overline{x} = \infty$ , Lemma 7 implies  $G^{-1}(1-t) \in RV_{-\gamma}^0$  where  $\gamma$  is the tail index of G. Therefore, we have  $\lim_{t\to 0} \frac{G^{-1}(1-at)}{G^{-1}(1-t)} = a^{-\gamma}$  and hence we obtain  $\lim_{t\to 0} \frac{\phi(G^{-1}(1-at))}{\phi(G^{-1}(1-t))} = \lim_{t\to 0} \left(\frac{G^{-1}(1-at)}{G^{-1}(1-t)}\right)^{\beta} = a^{-\beta\gamma}$ , so we have  $\phi(G^{-1}(1-t)) \in RV_{\rho}^0$  where  $\rho = -\gamma\beta$ . Also, we have  $\rho > -1$  if and only if  $-\gamma\beta > -1$  or  $\beta < 1/\gamma$ , which we assume. Finally, parts (ii) and (iii) follow directly from Corollary 4. Q.E.D.

#### **APPENDIX E: INVARIANT SEARCH TECHNOLOGIES**

A search technology  $P_n$  is *invariant*, as defined in Lester et al. (2015), if and only if Definition 3 holds.<sup>36</sup> In Lemma 1, we showed that mixed Poisson search technologies satisfy the key property (19) in Definition 3. Also, by Lemma 4, we know that  $P_0$  is continuous and infinitely differentiable for any  $\theta > 0$ . Therefore, any mixed Poisson search technology is invariant.

DEFINITION 3—Invariance: A search technology  $P_n$  is *invariant* if, for all  $y \in [0, 1]$ ,

$$\sum_{n=0}^{\infty} P_n(\theta) y^n = P_0(\theta(1-y))$$
(19)

where  $\mathbb{E}_P[N(\theta)] = \theta$  and  $P_0 : \mathbb{R}^+ \to [0, 1]$  is continuous and infinitely differentiable for  $\theta > 0$ .

Remarkably, the opposite is also true: any invariant search technology can be represented as a mixed Poisson distribution for some distribution F with  $\mathbb{E}_F[X] = 1$ . This result is due to Cai et al. (2025). We use Lemma 9 to derive this result here.

LEMMA 9—Laplace Transform: If  $P_n$  is an invariant search technology with mean  $\theta$ ,

$$P_0(\theta) = \int_0^\infty e^{-\theta\tau} dF(\tau)$$

for some probability distribution with cdf F and mean  $\mathbb{E}_F[X] = 1$ .

PROOF: If  $P_n$  is invariant, then  $P_0$  is continuous and infinitely differentiable. Moreover, as shown in Cai et al. (2025), we have  $(-1)^k P_0^{(k)}(z) \ge 0$  for all  $k \in \mathbb{N}$  and  $z \in \mathbb{R}^+$  where  $P_0^{(k)}(z)$  is the k-th derivative of  $P_0(z)$ . Therefore,  $P_0$  is a completely monotone function. By the Bernstein-Widder theorem, there exists a Laplace transform representation given by  $P_0(\theta) = \int e^{-\theta \tau} dF(\tau)$  for some finite measure F on  $\mathbb{R}^+$ . Because  $P_0(0) = 1$  by Definition 3, it follows that F must be a probability measure. Also, the mean of F equals one. To see this, we have the following expression for the k-th derivative:

$$P_0^{(k)}(\theta) = (-1)^k \int \tau^k e^{-\theta\tau} dF(\tau).$$
 (20)

Setting k = 1, we obtain  $P'_0(0) = -\int \tau dF(\tau)$ . Finally,  $P'_0(0) = -1$  so  $\int \tau dF(\tau) = 1$ . Q.E.D.

The following corollary of Lemma 9 is again due to Cai et al. (2025).

COROLLARY 6—Invariance implies Mixed Poisson: If  $P_n$  is an invariant search technology, then it is a mixed Poisson distribution for some distribution with cdf F and  $\mathbb{E}_F(X) = 1$ .

**PROOF:** From Lester et al. (2015), we know  $P_n$  satisfies Definition 3 if and only if

$$P_n(\theta) = \frac{(-1)^n \theta^n P_0^{(n)}(\theta)}{n!}.$$

Using (20) for the k-th derivative, and setting k = n and  $\theta = 0$ , we obtain (1). Q.E.D.

<sup>&</sup>lt;sup>36</sup>For a discussion of the intuition behind the "invariance" property, see Lester et al. (2015).

$e^{-z}$
$\left(\frac{r}{+z}\right)^r$
$\frac{1}{1+z}$
$\frac{e^{-\underline{\tau}z}}{\overline{z}+z}$
$\frac{-e^{-z}}{z}$
$\frac{1}{\overline{\tau} - e^{-\overline{\tau}z}}$
$\Gamma(-\alpha, \underline{\tau}z)$

TABLE II Examples of Mixed Poisson Search Technologies and their  $P_0$  Functions

Note: The distributions F listed in this table do not necessarily have means equal to one.

# APPENDIX F: DYNAMIC APPLICATIONS

Suppose there is a distribution F of agent types. Agent types are permanent. Assume the number of draws received by an agent of type  $\tau$  at time  $t \in \{1, 2, ..., T\}$  is a Poisson random variable with parameter  $\tau \theta_t$ . If we assume  $\mathbb{E}_F[X] = 1$ , Lemma 10 tells us that the cumulative number of draws for any agent at time T is given by a mixed Poisson distribution with mean  $\hat{\theta}_T = \sum_{t=1}^T \theta_t$  and the same distribution of types F.<sup>37</sup> In dynamic settings where agents can accumulate draws over time, we can therefore apply our asymptotic results in the limit as  $T \to \infty$  provided that  $\hat{\theta}_T \to \infty$  in this limit.<sup>38</sup>

LEMMA 10—Mixed Poisson Result for Dynamic Applications: Suppose agents have permanent types  $\tau$  drawn from a distribution F with  $\mathbb{E}_F[X] = 1$ . If the number of draws an agent of type  $\tau$  receives from the underlying distribution G at time  $t \in \{1, 2, ..., T\}$  is a Poisson random variable with mean  $\tau \theta_t$  where  $\theta_t \in \mathbb{R}_+$ , the random variable  $\hat{N}_T$  equal to the total number of draws an agent receives during periods  $t \in \{1, 2, ..., T\}$  is a mixed Poisson random variable with mean  $\hat{\theta}_T \equiv \sum_{t=1}^T \theta_t$  and mixing distribution equal to the type distribution F.

PROOF: Suppose the number of draws  $N_t^{\tau}$  received by an agent of type  $\tau$  in period  $t \in \{1, 2, ..., T\}$  is a Poisson random variable with parameter  $\tau \theta_t$ . Consider the total number of

<sup>&</sup>lt;sup>37</sup>The result still holds if  $\mathbb{E}_F[X] = \mu$ , but the mixed Poisson distribution has mean  $\mu \hat{\theta}_T$ .

<sup>&</sup>lt;sup>38</sup>In particular, we obtain  $\hat{\theta}_T \to \infty$  and we can apply our asymptotic results in the limit as  $T \to \infty$  if and only if  $\lim_{T\to\infty} \hat{\theta}_T = \lim_{T\to\infty} \sum_{t=1}^T \theta_t = +\infty$ . For example,  $\theta_t = \theta$  for all t.

draws  $\hat{N}_T^{\tau} \equiv \sum_{t=1}^T N_t^{\tau}$  received by an agent of type  $\tau$  in any period up to and including T. It is well known that the finite sum of Poisson random variables is Poisson with mean equal to the sum of the means. Therefore, we know that  $\hat{N}_T^{\tau}$  is a Poisson random variable with mean  $\tau \hat{\theta}_T$  where  $\hat{\theta}_T \equiv \sum_{t=1}^T \theta_t$ . Now let  $N_t$  denote the number of draws received by an agent in period  $t \in \{1, 2, \dots, T\}$  and define  $\hat{N}_T \equiv \sum_{t=1}^T N_t$ . The distribution of types  $\tau$  has cdf F, so  $\Pr(\hat{N}_T = 0) = \int e^{-\tau \hat{\theta}_T} dF(\tau)$ . It follows that Assumption 2 holds and  $\hat{N}_T$  is a mixed Poisson random variable with mean  $\hat{\theta}_T$  and mixing distribution F. Q.E.D.

### APPENDIX G: GENERALIZATION OF RESULT IN JONES (2023)

We adopt the approach in Jones (2023) to derive a general result regarding the asymptotic behavior of the maximum for any underlying distribution. We then use this result to derive the extreme value distribution for the Pareto example. We first present the analogous result in Jones (2023) and then show how this result generalizes to our environment.

#### G.1. Fixed Number of Draws

Let  $X_1, \ldots, X_n$  be i.i.d. draws from an underlying distribution G where  $n \ge 1$ . Defining the random variable  $M_n \equiv \max\{X_1, \ldots, X_n\}$ , we have  $\Pr(M_n \le x) = G(x)^n$ .

Now define a new random variable,  $\hat{M}_n \equiv n(1 - G(M_n))$ . As Jones (2023) shows,

$$\Pr(\hat{M}_n \le y) = 1 - \left(1 - \frac{y}{n}\right)^n.$$
(21)

Taking the limit as  $n \to \infty$  delivers the result in Jones' Theorem 1,

$$\Pr(\hat{M}_n \ge y) = e^{-y}.$$

Suppose that G is Pareto, i.e.  $G(x) = 1 - x^{-1/\gamma}$ . Because  $\hat{M}_n$  is asymptotically exponentially distributed, we have  $n(1 - G(M_n)) = \varepsilon + o_p(1)$  where  $\varepsilon$  is a random variable with  $df 1 - e^{-y}$ . Therefore,  $M_n = n^{\gamma} (\varepsilon + o_p(1))^{-\gamma}$  and, for n large,  $M_n \approx n^{\gamma} \varepsilon^{-\gamma}$ . Defining  $\tilde{\varepsilon} \equiv \varepsilon^{-\gamma}$ , it is straightforward to verify that  $\Pr(\tilde{\varepsilon} \leq x) = e^{-x^{-1/\gamma}}$ . We thus obtain the well-known result that the extreme value distribution is Fréchet:

$$H_{\gamma}(x) = e^{-x^{-1/\gamma}}$$

### G.2. Random Number of Draws

Suppose the number of draws is a random variable  $N(\theta)$  with mean  $\theta$  and mixed Poisson distribution  $P_n$ . Define a random variable,  $M_{N(\theta)} \equiv \max\{X_1, \ldots, X_{N(\theta)}\}$ . For consistency with Jones (2023), this is defined only for  $N(\theta) \ge 1$ . Using Lemma 3, we obtain

$$\Pr(M_{N(\theta)} \le x) = \frac{P_0(\theta(1 - G(x))) - P_0(\theta)}{1 - P_0(\theta)}.$$

Now define a new random variable,  $\hat{M}_{N(\theta)} \equiv \theta(1 - G(M_{N(\theta)}))$ . Analogously to (21),

$$\Pr(\hat{M}_{N(\theta)} \le y) = 1 - \Pr\left(M_{N(\theta)} \le G^{-1}\left(1 - \frac{y}{\theta}\right)\right).$$

Combining these, we obtain Theorem 4, which generalizes Corollary 1 in Jones (2023).

THEOREM 4: Suppose  $N(\theta)$  is a random variable with mixed Poisson distribution  $P_n$  and mean  $\theta$ . Let  $M_{N(\theta)}$  denote the maximum value from  $N(\theta) \ge 1$  independent draws from a distribution G and define  $\hat{M}_{N(\theta)} \equiv \theta(1 - G(M_{N(\theta)}))$ . For  $y \ge 0$ , we have

$$\Pr(\hat{M}_{N(\theta)} \ge y) = \frac{P_0(y) - P_0(\theta)}{1 - P_0(\theta)}.$$

In the limit as  $\theta \to \infty$ , we have  $\lim_{\theta \to \infty} \Pr(\hat{M}_{N(\theta)} \ge y) = P_0(y)$ .

Suppose that G is Pareto, i.e.  $G(x) = 1 - x^{-1/\gamma}$ . For large  $\theta$ , Theorem 4 implies  $\hat{M}_{N(\theta)} = \varepsilon_P + o_p(1)$  where  $\varepsilon_P$  is a random variable with cdf  $1 - P_0(y)$ . Therefore,  $M_{N(\theta)} = \theta^{\gamma}(\varepsilon_P + o_p(1))^{-\gamma}$  and, for large  $\theta$ , we have  $M_{N(\theta)} \approx \theta^{\gamma} \varepsilon_P^{-\gamma}$ . Defining  $\tilde{\varepsilon}_P \equiv \varepsilon_P^{-\gamma}$ , we have  $\Pr(\tilde{\varepsilon}_P \leq x) = P_0(x^{-1/\gamma})$ . We obtain the following extreme value distribution:

$$H_{\gamma,P}(x) = P_0(x^{-1/\gamma}).$$
(22)

Theorem 4 thus delivers an alternative derivation of expression (6) in Section 3.<sup>39</sup>

#### REFERENCES

- ADELL, JOSÉ AND JESÚS DE LA CAL (1993): "On the Uniform Convergence of Normalized Poisson Mixtures to their Mixing Distribution," *Statistics & Probability Letters*, 18 (3), 227–232. [29]
- ARANDELOVIC, D (1976): "Sur un Theoreme Mercerien Asymptotique General pour des Fonctions a Comportement Regulier," Publ. Inst. Math., Nouv. Sér, 20 (34), 29–36. [33]
- BARNDORFF-NIELSEN, OLE (1964): "On the Limit Distribution of the Maximum of a Random Number of Independent Random Variables," Acta Math. Acad. Sci. Hungar, 15, 399–403. [28]
- BEARE, BRENDAN K AND ALEXIS AKIRA TODA (2022): "Determination of Pareto exponents in Economic Models Driven by Markov Multiplicative Processes," *Econometrica*, 90 (4), 1811–1833. [2]
- BENHABIB, JESS AND ALBERTO BISIN (2018): "Skewed Wealth Distributions: Theory and Empirics," Journal of Economic Literature, 56 (4), 1261–1291. [2]
- BINGHAM, N. H., C. M. GOLDIE, AND J. L. TEUGELS (1987): *Regular Variation*, Cambridge University Press. [14, 33, 34]
- BOEHM, JOHANNES AND EZRA OBERFIELD (2020): "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *The Quarterly Journal of Economics*, 135 (4), 2007–2058. [3]

(2023): "Growth and the Fragmentation of Production," *Working paper*. [3]

BOUCHER, VINCENT, MICHELLE RENDALL, PHILIP USHCHEV, AND YVES ZENOU (2024): "Toward a General Theory of Peer Effects," *Econometrica*, 92 (2), 543–565. [24]

- BUERA, FRANCISCO J. AND ROBERT E. LUCAS (2018): "Idea Flows and Economic Growth," Annual Review of Economics, 10 (1), 315–345. [2]
- BUERA, FRANCISCO J. AND EZRA OBERFIELD (2020): "The Global Diffusion of Ideas," *Econometrica*, 88 (1), 83–114. [2]
- CAI, XIAOMING, PIETER GAUTIER, AND RONALD WOLTHOFF (2025): "Spatial Search," Journal of Economic Theory, 105976. [2, 5, 37]
- CAMPBELL, ARTHUR, PHILIP USHCHEV, AND YVES ZENOU (2024): "The Network Origins of Entry," Journal of Political Economy, 132 (11), 3867–3916. [2, 25]
- DE HAAN, LAURENS AND ANA FERREIRA (2006): *Extreme Value Theory: An Introduction*, Springer, New York. [12, 29, 31]
- EATON, JONATHAN AND SAMUEL KORTUM (1999): "International Technology Diffusion: Theory and Measurement," International Economic Review, 40 (3), 537–570. [2]

<sup>&</sup>lt;sup>39</sup>Another approach which works nicely for the Pareto distribution is to start with the distribution of the maximum  $M_{N(\theta)}$  for *finite*  $\theta$  given by Lemma 3, i.e.  $H_P(x;\theta) = P_0(\theta(1 - G(x)))$ . If the Pareto distribution is  $G(x) = 1 - (x/\underline{x})^{-1/\gamma}$  with lower bound  $\underline{x}$ , then for  $x \ge \underline{x}$  we have  $H_P(x;\theta) = P_0(\theta(x/\underline{x})^{-1/\gamma})$ . If we let  $\theta \to \infty$  and  $\underline{x} \to 0$ , holding constant  $K = \theta \underline{x}^{1/\gamma}$ , we get  $P_0(Kx^{-1/\gamma})$ . If we define  $Z_{N(\theta)} \equiv M_{N(\theta)}/K^{\gamma}$ , we get the distribution (22). We thank an anonymous referee for suggesting this alternative approach.

- EECKHOUT, JAN AND PHILIP KIRCHER (2010): "Sorting Versus Screening: Search Frictions and Competing Mechanisms," *Journal of Economic Theory*, 145 (4), 1354–1385. [5]
- GABAIX, XAVIER (1999): "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, 739–767. [2]
  - (2009): "Power Laws in Economics and Finance," Annual Review of Economics, 1, 255–293. [2]
- (2016): "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30 (1), 185–206.
- GABAIX, XAVIER, DAVID LAIBSON, DEYUAN LI, HONGYI LI, SIDNEY RESNICK, AND CASPER G. DE VRIES (2016): "The Impact of Competition on Prices with Numerous Firms," *Journal of Economic Theory*, 165 (C), 1–24. [2, 4, 14, 15, 19, 21, 22, 23, 26, 34]
- GALAMBOS, JÁNOS (1987): *The Asymptotic Theory of Extreme Order Statistics*, New York: John Wiley & Sons. [28]
- HART, OLIVER D (1985): "Monopolistic Competition in the Spirit of Chamberlin: Special Results," *The Economic Journal*, 95 (380), 889–908. [21]
- HAUSMAN, JERRY A, BRONWYN H HALL, AND ZVI GRILICHES (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52 (4), 909–938. [2, 18]
- JOHNSON, NORMAN L, ADRIENNE W KEMP, AND SAMUEL KOTZ (2005): Univariate Discrete Distributions, vol. 444, John Wiley & Sons. [6]
- JONES, CHARLES I. (2005): "The Shape of Production Functions and the Direction of Technical Change," *Quarterly Journal of Economics*, 120 (2), 517–549. [2, 3]
- JONES, CHARLES I (2023): "Recipes and Economic Growth: A Combinatorial March down an Exponential Tail," Journal of Political Economy, 131 (8), 1994–2031. [2, 3, 9, 39]
- JONES, CHARLES I AND JIHEE KIM (2018): "A Schumpeterian Model of Top Income Inequality," *Journal of Political Economy*, 126 (5), 1785–1826. [2]
- KARLIS, DIMITRIS AND EVDOKIA XEKALAKI (2005): "Mixed Poisson Distributions," International Statistical Review/Revue Internationale de Statistique, 35–58. [6, 7]
- KORTUM, SAMUEL (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65 (6), 1389–1419. [2, 3]
- KUBA, MARKUS AND ALOIS PANHOLZER (2016): "On Moment Sequences and Mixed Poisson Distributions," Probability Surveys, 13, 89–155. [29]
- LESTER, BENJAMIN, LUDO VISSCHERS, AND RONALD WOLTHOFF (2015): "Meeting Technologies and Optimal Trading Mechanisms in Competitive Search Markets," *Journal of Economic Theory*, 155, 1–15. [2, 5, 37]
- LUCAS, ROBERT E. (2009): "Ideas and Growth," *Economica*, 76 (301), 1–19. [2]
- LUTTMER, ERZO GJ (2007): "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 1103–1144. [2]
- MANGIN, SEPHORAH (2017): "A Theory of Production, Matching, and Distribution," *Journal of Economic Theory*, 172, 376–409. [2, 3]
- ——— (2024): "When is Competition Price-increasing? The Impact of Expected Competition on Prices," *The RAND Journal of Economics*, 55 (4), 627–657. [2, 15, 16, 22, 31, 32]
- MANGIN, SEPHORAH AND PETR SEDLÁČEK (2018): "Unemployment and the Labor Share," *Journal of Monetary Economics*, 94, 41–59. [3]
- MARTELLINI, PAOLO AND GUIDO MENZIO (2020): "Declining Search Frictions, Unemployment, and Growth," Journal of Political Economy, 128 (12), 4387–4437. [2]
- MELITZ, MARC J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71 (6), 1695–1725. [17]
- NEWMAN, MARK EJ, STEVEN H STROGATZ, AND DUNCAN J WATTS (2001): "Random Graphs with Arbitrary Degree Distributions and their Applications," *Physical Review E*, 64 (2), 026118. [24]
- OBERFIELD, EZRA (2018): "A Theory of Input–Output Architecture," Econometrica, 86 (2), 559–589. [2, 3]
- PERLOFF, JEFFREY M AND STEVEN C SALOP (1985): "Equilibrium with Product Differentiation," *The Review of Economic Studies*, 52 (1), 107–120. [21]
- RESNICK, SIDNEY I. (1987): Extreme Values, Regular Variation, and Point Processes, Springer. [14]
- RHODES, ANDREW AND JIDONG ZHOU (2024): "Personalized Pricing and Competition," American Economic Review, 114 (7), 2141–70. [21]
- SATTINGER, MICHAEL (1984): "Value of an Additional Firm in Monopolistic Competition," *The Review of Economic Studies*, 51 (2), 321–332. [21]
- SCHILLING, RENÉ L, RENMING SONG, AND ZORAN VONDRACEK (2012): Bernstein Functions: Theory and Applications, De Gruyter. [28]
- SMITH, RICHARD L (1984): "Threshold Methods for Sample Extremes," in *Statistical Extremes and Applications*, Springer, 621–638. [3]