The Effect of Search Frictions on Extreme Outcomes*

Louis Becker† and Sephorah Mangin‡

November 28, 2022

Preliminary and incomplete

Abstract

Extreme value processes involving maxima are widespread in economics. We provide some general results regarding the asymptotic effect of search frictions on the outcomes of such processes. To do this, we allow the number of draws from the *underlying distribution* (e.g. of productivities, efficiencies, or ideas) to be given by a discrete probability distribution called the *search technology.* We show that extreme value outcomes depend not only on the underlying distribution and its tail index, but also on the search technology. For example, if the underlying distribution is Pareto, the extreme value distribution is Fréchet if and only if the search technology is either Poisson or degenerate. We consider some applications of our results to both aggregate productivity and markups.

*We thank Xiaoming Cai, Michael Howes, Hongyi Li, George Mailath, and Ronald Wolthoff for useful discussions.

†Research School of Economics, Australian National University

‡Research School of Economics, Australian National University

1 Introduction

Extreme value processes involving maxima 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, this process involves taking the maximum of a number of draws from a distribution (e.g. of productivities, efficiencies, ideas, or utility shocks) and then considering the behavior of the maximum in the limit as the number of draws becomes large. It is widely known that the limiting distribution of the maximum (appropriately normalized) must be either Fréchet, Gumbel, or Weibull. For example, fat-tailed distributions such as the widely-used Pareto distribution give rise to the Fréchet extreme value distribution. Variations of this type of Pareto-Fréchet extreme value process are used in a large class of models in macroeconomics and trade, including Kortum (1997), Eaton and Kortum (1999), Jones (2005), Lucas (2009), Buera and Lucas (2018), Oberfield (2018), Buera and Oberfield (2020), and Jones (2022).

In this paper, we generalize this type of extreme value process by allowing the number of draws from the *underlying distribution* (e.g. of productivities, efficiencies, ideas, or utility shocks) to be given by a discrete probability distribution called the *search technology.* After incorporating this new source of randomness, we then consider the limit as the expected number of draws from the underlying distribution becomes large. While we interpret this randomness as arising from search frictions that lead to dispersion in the number of draws, this dispersion could potentially result from many other sources. Regardless of the source, we can interpret this dispersion either as reflecting cross-sectional dispersion across a large number of firms or agents, or as reflecting uncertainty faced by a single firm or agent regarding the number of draws.

This paper provides two main theorems that are quite general and may have broader application beyond the problems we consider here. For both results, we restrict attention to a class of search technologies called *invariant* in Lester, Visschers, and Wolthoff (2015). 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. This second result provides a generalization of an analogous result in Gabaix, Laibson, Li, Li, Resnick, and de Vries (2016) where the number of draws is deterministic. We use both of our general results to examine the asymptotic effect of search frictions on extreme value outcomes.

We might expect that in the limit as the expected number of draws becomes large, the effect of search frictions would disappear and that the distribution of the maximum would eventually behave in a standard way. That is, we might expect this limit to be "frictionless" in some sense. Surprisingly, however, we find that search frictions still matter – even asymptotically. We show that the search technology can affect the asymptotic behavior of the maximum, or any functions of the maximum, and it may thereby affect important economic outcomes such as productivity or markups.

Remarkably, the extreme value distribution need not be any of the three standard types of distributions – Fréchet, Gumbel, and Weibull. Instead, the extreme value distribution depends not only on the underlying distribution and its tail index, but also on the nature of the *search technology*. For example, if the distribution of productivities is Pareto, the extreme value distribution is not necessarily Fréchet, and the expected value of the maximum does not necessarily behave like the expected value of the Fréchet distribution. Instead, the expected value of the maximum behaves asymptotically like the expected value of the Fréchet distribution scaled by a new term capturing the effect of search frictions. This scaling factor depends on both the tail index of the distribution of productivities and on the search technology. Depending on the details of the application, we find that extreme value outcomes may be either increasing or decreasing in the asymptotic dispersion of the search technology.

We show that the extreme value distribution takes the standard form if and only if the search technology is either Poisson or degenerate, i.e. there is zero asymptotic dispersion. This means that there is no asymptotic effect of search frictions if and only if the search technology is Poisson or degenerate. The Poisson distribution can therefore be viewed as an extreme special case where the effect of search frictions disappears asymptotically. This explains why the Poisson distribution can deliver the standard extreme value results, a fact which partly explains why this distribution is often used in the literature using extreme value theory in economics. For example, the Poisson distribution is used in Kortum (1997), Jones (2005), Oberfield (2018), Boehm and Oberfield (2020), Boehm and Oberfield (2022), and Jones (2022).

We consider two applications of our results: aggregate productivity and markups. We illustrate our results for both applications by using a class of search technologies, the negative binomial family, that delivers tractable expressions. This class nests the Geometric distribution as a special case and the widely-used Poisson distribution as a limiting case. It may be useful for empirical applications because it features a single parameter which governs the effect of search frictions on extreme outcomes.

First, we consider an application of our results to the level of aggregate productivity and the cross-sectional distribution of firm productivity. In general, we find that greater asymptotic dispersion in the number of productivity draws *decreases* aggregate productivity if the tail index of the underlying distribution of productivities is positive. If the tail index is zero, there is no asymptotic effect on aggregate productivity.

We can think of asymptotic dispersion as a kind of misallocation that reduces output compared to the frictionless economy where the number of draws is not random. We show how to quantify the productivity loss from search frictions and find that it can be significant. For example, a change in the search technology alone from Poisson to Geometric (keeping the expected number of firm productivity draws constant) can decrease aggregate productivity by around 11%, while also increasing cross-sectional productivity dispersion (as measured by the coefficient of variation) by 28%. This suggests that search frictions are of first-order importance not only for reducing unemployment (as is well-known) but potentially also for aggregate output. By abstracting from the role of search frictions in models that use extreme value processes in both macroeconomics and trade, we ignore these important effects of the search technology.

Second, we consider an application of our results to the asymptotic behavior of markups in a discrete choice model with random utility shocks. In the setting we consider, prices are determined by limit pricing (or "personalized pricing"), sometimes referred to as Bertrand competition. We apply our generalization of the result in Gabaix et al. (2016) which describes the asymptotic behavior of functions of the maximum. In general, we find that the search technology can have a significant effect on asymptotic markups – either positively or negatively – depending on the tail index of the underlying distibution of utility shocks. Greater asymptotic dispersion in the number of productivity draws *decreases* markups if the tail index is positive, but it *increases* asymptotic markups if the tail index is negative.

Our results on markups also apply more generally to the asymptotic behavior of buyer surplus and seller revenue in auctions with a large number of bidders. This is because the markup equals the difference between the highest and second-highest utility shock, which is equal to the buyer surplus in a second-price auction.

2 Preliminaries

Suppose there are $n \in \mathbb{N}$ draws from a distribution of values. We assume that the distribution of values has cdf G and it satisfies **A1**. We call the distribution G the *underlying distribution*. Depending on the specific application, it may be a distribution of productivities, ideas, efficiencies, or utility shocks.

Assumption 1 (**A1**)**.** *The distribution of values x has a twice-differentiable cdf G with pdf $g = G' > 0$, a finite mean, and support $[\underline{x}, \overline{x}] \subseteq \mathbb{R}$ where $\underline{x}, \overline{x} \in \mathbb{R} \cup \{\pm\infty\}$.*[1]

We assume the number of draws is itself a random variable that has a discrete probability distribution with pmf $\mathbb{P}_n$ and support $\mathbb{N}$. For any $n \in \mathbb{N}$, the probability there are n draws is given by $\mathbb{P}_n(\theta)$ where $\theta \in \mathbb{R}_+$ is the expected number of draws. We denote by $N(\theta)$ the random variable with distribution $\mathbb{P}_n$. We assume that $\lim_{\theta\to\infty} \mathbb{P}_0(\theta) = 0$. We call the distribution $\mathbb{P}_n$ the *search technology*.

Observe that there are two potential sources of randomness: the value of each draw x, and the number n of draws. The first source of randomness is the *underlying distribution* G, and the second source of randomness is the *search technology* $\mathbb{P}_n$.

It is useful to define the probability generating function of the distribution $\mathbb{P}_n$.

Definition 1. *The* probability generating function *of $\mathbb{P}_n(\theta)$ is given by*

$$\mathbb{E}_{\mathbb{P}}(y^n) = \sum_{n=0}^{\infty} \mathbb{P}_n(\theta) y^n. \tag{1}$$

We assume that the search technology $\mathbb{P}_n$ is *invariant*, as defined in Lester et al. (2015). This means that the probability generating function can be written as a function of a single variable. This function must be equal to $\mathbb{P}_0$.

Assumption 2 (**A2**)**.** *The distribution $\mathbb{P}_n(\theta)$ is* invariant*, i.e. for all $y \in [0, 1]$,*

$$\sum_{n=0}^{\infty} \mathbb{P}_n(\theta) y^n = \mathbb{P}_0(\theta(1 - y)) \tag{2}$$

where $\mathbb{P}_0 : \mathbb{R}^+ \to [0, 1]$ is continuous and infinitely differentiable.

[1] By a slight abuse of notation, if $\overline{x} = +\infty$ then we intend $[\underline{x}, \overline{x}]$ to mean $[\underline{x}, +\infty)$ and if $\underline{x} = -\infty$ then we intend $[\underline{x}, \overline{x}]$ to mean $(-\infty, \overline{x}]$.

The function $\mathbb{P}_0$ will turn out to be extremely useful as it will capture everything we need to know about the search technology. For future reference, it is convenient to summarize some properties of the function $\mathbb{P}_0$ here.

Lemma 1. *If $\mathbb{P}_n$ satisfies* **A2***, the function $\mathbb{P}_0 : \mathbb{R}^+ \to [0,1]$ has the following properties: (i) $\mathbb{P}_0(0) = 1$; (ii) $\mathbb{P}_0'(0) = -1$; and (iii) $\lim_{\theta\to\infty} \mathbb{P}_0(\theta) = 0$.*

We now present a couple of useful examples of invariant search technologies.

Example 1: If $\mathbb{P}_n$ is a Poisson search technology, then

$$\mathbb{P}_n(\theta) = \frac{e^{-\theta}\theta^n}{n!} \tag{3}$$

and the function $\mathbb{P}_0$ is

$$\mathbb{P}_0(z) = e^{-z} \tag{4}$$

and the probability generating function is

$$\sum_{n=0}^{\infty} \mathbb{P}_n(\theta) y^n = e^{-\theta(1-y)}. \tag{5}$$

So, the Poisson search technology satisfies **A2**.

Example 2: If $\mathbb{P}_n$ is a Geometric search technology, then

$$\mathbb{P}_n(\theta) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^n \tag{6}$$

and the function $\mathbb{P}_0$ is

$$\mathbb{P}_0(z) = \frac{1}{1+z} \tag{7}$$

and the probability generating function is

$$\sum_{n=0}^{\infty} \mathbb{P}_n(\theta) y^n = \frac{1}{1+\theta(1-y)}. \tag{8}$$

So, the Poisson search technology satisfies **A2**.

We are interested in obtaining extreme value results regarding the behavior of the distribution of the maximum in the limit as θ goes to infinity. In Section 5, we derive the general form of the extreme value distribution for any search technology that is invariant. In Section 6, we derive a general expression regarding the asymptotic behavior of extreme value outcomes, i.e. functions of the maximum, for any search technology that is invariant. Before turning to these general methodological results, we first provide a glimpse into the general results we will obtain by adopting the approach in Jones (2022). In Section 4, we then describe some useful properties of invariant search technologies that will help us derive our general results.

3 Preview of results

In this section, we follow the approach in Jones (2022). We first derive a result regarding the asymptotic behavior of the maximum and then use this result to provide a simple heuristic derivation of the extreme value distribution for a specific example. This will give us a preview of the general result we obtain in Section 5. In particular, we will see clearly how the extreme value distribution depends not only on the underlying distribution and its tail index, but also on nature of the search technology.

First, we briefly reproduce the approach in Jones (2022) and then show how it generalizes to our environment where the number of draws is random.

Let $X_1, ..., X_n$ be i.i.d. random variables with distribution G. Define a random variable for the maximum, $M_n \equiv \max\{X_1, ..., X_n\}$. It is well-known that the cdf of M_n for $n \geq 1$ is $\Pr(M_n \leq x) = G(x)^n$. Now define a new random variable, $\hat{M}_n \equiv (1 - G(M_n))$. As Jones (2022) shows, this implies

$$\Pr(\hat{M}_n \leq x) = 1 - \left(1 - \frac{x}{n}\right)^n. \tag{9}$$

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

$$\Pr(\hat{M}_n \leq x) = 1 - e^{-x}. \tag{10}$$

That is, the random variable $\hat{M}_n$ is exponentially distributed as $n \to \infty$.

Now suppose the underlying distribution is Pareto, i.e. $G(x) = 1 - x^{-1/\gamma}$. Because $\hat{M}_n$ is exponentially distributed by Jones' Theorem 1, we have $n(1-G(M_n)) = \varepsilon + o_p(1)$ where ε is an exponentially distributed random variable with cdf $1 - e^{-x}$. Therefore,

$M_n = n^\gamma(\varepsilon + o_p(1))^{-\gamma}$ and, for n large, we have $M_n \approx n^\gamma \varepsilon^{-\gamma}$. Defining $\tilde{\varepsilon} \equiv \varepsilon^{-1/\gamma}$, it is straightforward to verify that $\Pr(\tilde{\varepsilon}_{\mathbb{P}} \leq x) = e^{-x^{-1/\gamma}}$ and we obtain the well-known result that the extreme value distribution is Fréchet:

$$H_\gamma(x) = e^{-x^{-1/\gamma}}. \tag{11}$$

Now, we adapt Jones' argument above to our setting. Suppose that $N(\theta)$ is a random variable with mean θ and distribution $\mathbb{P}_n$. Define a random variable for the maximum, $M_{N(\theta)} \equiv \max\{X_1, ..., X_{N(\theta)}\}$. Defining the cdf of the maximum by $H_{\mathbb{P}}(x;\theta) = \Pr(M_{N(\theta)} \leq x \mid N(\theta) \geq 1)$, we obtain

$$H_{\mathbb{P}}(x;\theta) = \frac{\sum_{n=0}^{\infty} \mathbb{P}_n(\theta) G(x)^n - \mathbb{P}_0(\theta)}{1 - \mathbb{P}_0(\theta)}. \tag{12}$$

Therefore, if $\mathbb{P}_n$ satisfies **A2**, we obtain:

$$H_{\mathbb{P}}(x;\theta) = \frac{\mathbb{P}_0(\theta(1 - G(x))) - \mathbb{P}_0(\theta)}{1 - \mathbb{P}_0(\theta)}. \tag{13}$$

We can now provide a generalization of Theorem 1 and Corollary 1 in Jones (2022). Define a new random variable, $\hat{M}_{N(\theta)} \equiv \theta(1 - G(M_{N(\theta)}))$. Analogous to (9), we have

$$\Pr(\hat{M}_{N(\theta)} \leq x) = 1 - \Pr(M_{N(\theta)} \leq G^{-1}(1 - x/\theta)). \tag{14}$$

Using expression (13) for $H_{\mathbb{P}}(x;\theta)$, we obtain[2]

$$\Pr(\hat{M}_{N(\theta)} \leq x) = \frac{1 - \mathbb{P}_0(x)}{1 - \mathbb{P}_0(\theta)}. \tag{15}$$

In the limit as $\theta \to \infty$, we have $\lim_{\theta \to \infty} \mathbb{P}_0(\theta)$ so

$$\Pr(\hat{M}_{N(\theta)} \leq x) = 1 - \mathbb{P}_0(x). \tag{16}$$

The random variable $\hat{M}_{N(\theta)}$ is no longer exponentially distributed as $\theta \to \infty$, except

[2]To see this, note that $\Pr(\hat{M}_{N(\theta)} \leq x) = 1 - \Pr(M_{N(\theta)} \leq G^{-1}(1 - x/\theta))$, which is equal to

$$1 - \frac{\mathbb{P}_0(\theta(1 - G(G^{-1}(1 - x/\theta)))) - \mathbb{P}_0(\theta)}{1 - \mathbb{P}_0(\theta)} = 1 - \frac{\mathbb{P}_0(x) - \mathbb{P}_0(\theta)}{1 - \mathbb{P}_0(\theta)}$$

in the special case where $\mathbb{P}_0(x) = e^{-x}$ and the search technology is Poisson. This will turn out to be crucial because we will later see that this result is very closely related to the form of the extreme value distribution.

We summarize this result in Lemma 2, which directly generalizes Theorem 1 in Jones (2022) to the case where $N(\theta)$ is a discrete random variable. Note that if $\mathbb{P}_n$ is degenerate, Theorem 1 in Jones (2022) follows from (17). If $\mathbb{P}_n$ is Poisson, Corollary 1 in Jones (2022) follows from (18) where $\mathbb{P}_0(x) = e^{-x}$.

Lemma 2. *Suppose that $N(\theta)$ is a random variable with mean θ and distribution $\mathbb{P}_n$. Let $M_{N(\theta)}$ denote the maximum value from $N(\theta) \geq 1$ independent draws from a distribution G that satisfies* **A1** *and define $\hat{M}_{N(\theta)} \equiv \theta(1 - G(M_{N(\theta)}))$. For $x \geq 0$,*

$$\Pr(\hat{M}_{N(\theta)} \geq x) = \frac{\sum_{n=1}^{\infty} \mathbb{P}_n(\theta)(1 - \frac{x}{\theta})^n}{1 - \mathbb{P}_0(\theta)}. \tag{17}$$

where $\mathbb{P}_0(\theta) \equiv \Pr(N(\theta) = 0)$. If $\mathbb{P}_n$ satisfies **A2***, then*

$$\Pr(\hat{M}_{N(\theta)} \geq x) = \frac{\mathbb{P}_0(x) - \mathbb{P}_0(\theta)}{1 - \mathbb{P}_0(\theta)} \tag{18}$$

and

$$\lim_{\theta \to \infty} \Pr(\hat{M}_{N(\theta)} \geq x) = \mathbb{P}_0(x). \tag{19}$$

With these results in hand, we now turn to the task of determining the extreme value distribution, i.e. the distribution of the (normalized) maximum in the limit as the expected number of draws θ goes to infinity. To do this, we adopt the simple heuristic argument used in Jones (2022). This approach works out nicely when the underlying distribution is Pareto. In general, however, we will need our result in Section 5 to derive an expression for the extreme value distribution.

Suppose the underlying distribution is Pareto, i.e. $G(x) = 1 - x^{-1/\gamma}$. Similarly to Jones' argument above, define a new random variable by $\hat{M}_{N(\theta)} \equiv \theta(1 - G(M_{N(\theta)})) = \theta M_{N(\theta)}^{-1/\gamma}$. Lemma 2 implies that $\theta(1 - G(M_{N(\theta)})) = \varepsilon_{\mathbb{P}} + o_p(1)$ where $\varepsilon_{\mathbb{P}}$ is a random variable with cdf $1 - \mathbb{P}_0(x)$. Therefore, we have $M_{N(\theta)} = \theta^{\gamma}(\varepsilon_{\mathbb{P}} + o_p(1))^{-\gamma}$ and, for θ large, $M_{N(\theta)} \approx \theta^{\gamma}\varepsilon_{\mathbb{P}}^{-\gamma}$. Defining $\tilde{\varepsilon}_{\mathbb{P}} \equiv \varepsilon_{\mathbb{P}}^{-1/\gamma}$, we have $\Pr(\tilde{\varepsilon}_{\mathbb{P}} \leq x) = \mathbb{P}_0(x^{-1/\gamma})$ and thus the extreme value distribution is given by

$$H_{\gamma,\mathbb{P}}(x) = \mathbb{P}_0(x^{-1/\gamma}). \tag{20}$$

The extreme value distribution $H_{\gamma,\mathbb{P}}(x)$ is no longer necessarily Fréchet.[3] In general, the extreme value distribution takes a form that depends on the search technology.

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, then $\mathbb{P}_0(z) = e^{-z}$ and the extreme value distribution is

$$H_{\gamma,\mathbb{P}}(x) = e^{-x^{-1/\gamma}} \tag{21}$$

if the distribution G is Pareto. That is, we obtain the standard Fréchet distribution.

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, then $\mathbb{P}_0(z) = \frac{1}{1+z}$ and the extreme value distribution is

$$H_{\gamma,\mathbb{P}}(x) = \frac{1}{1 + x^{-1/\gamma}}. \tag{22}$$

if the distribution G is Pareto. Observe that this differs from any of the three standard forms of the extreme value distribution (Fréchet, Gumbel, or Weibull).

After this preview of our results, we now turn to our general result. In Section 4, we first describe some useful properties of invariant technologies that will be used to prove some of our results. In Section 5, we then derive the general form of the extreme value distribution for any invariant search technology $\mathbb{P}_n$.

4 Properties of invariant search technologies

Before presenting our general results, we summarize some properties of invariant search technologies that will prove useful. First, we describe a useful representation of invariant search technologies. Next, we use this representation to present some results regarding the asymptotic behavior of invariant search technologies.

[3] To see this, $\Pr(\tilde{\varepsilon}_{\mathbb{P}} \leq x) = \Pr(\varepsilon_{\mathbb{P}}^{-1/\gamma} \leq x) = 1 - \Pr(\varepsilon_{\mathbb{P}} \leq x^{-1/\gamma}) = 1 - (1 - \mathbb{P}_0(x^{-1/\gamma})) = \mathbb{P}_0(x^{-1/\gamma})$

4.1 A useful representation

In this section, we describe a very useful representation of the class of invariant search technologies which is due to Cai, Gautier, and Wolthoff (2022). To obtain the results in this section, we adapt their approach to our environment.

Lemma 3 states that if $\mathbb{P}_n$ is invariant, then $\mathbb{P}_0$ can be expressed as the Laplace transform of some probability distribution. The idea of using Bernstein's theorem to obtain this representation is due to Cai et al. (2022).[4]

Lemma 3. *If $\mathbb{P}_n$ satisfies **A2**, then the function $\mathbb{P}_0$ can be written as*

$$\mathbb{P}_0(\theta) = \int e^{-\theta t} dF(t) \tag{23}$$

for some distribution with cdf F and $\mathbb{E}_F(x) = 1$.

The mixed Poisson class of distributions $\mathbb{P}_n$ is defined by

$$\mathbb{P}_n(\theta) = \int \frac{(\theta t)^n e^{-\theta t}}{n!} dF(t) \tag{24}$$

for some mixing distribution with cdf F and $\mathbb{E}_F(x) = 1$. Notice that $\mathbb{P}_0(\theta) = \int e^{-\theta t} dF(t)$, which is exactly the representation of invariant distributions that we derived in Lemma 3 above. Any mixed Poisson distribution is invariant, i.e. it satisfies **A2**, since

$$\begin{aligned}
\sum_{n=0}^{\infty} \mathbb{P}_n(\theta) y^n &= \sum_{n=0}^{\infty} y^n \int \frac{(\theta t)^n e^{-\theta t}}{n!} dF(t) \\
&= \int e^{-\theta(1-y)t} dF(t) \\
&= \mathbb{P}_0(\theta(1-y)).
\end{aligned}$$

A corollary of Lemma 3 is that any invariant distribution has a representation as a mixed Poisson distribution for some mixing distribution F. This result, summarized below, is due to Cai et al. (2022). In their environment, however, F is not necessarily a probability measure and Lemma 3 does not necessarily hold, so $\mathbb{P}_n$ is not necessarily a mixed Poisson distribution.

[4] In particular, Cai et al. (2022) apply Bernstein's theorem to the function $1 - \mathbb{P}_0(\theta)$, which they prove is a Bernstein function.

Corollary 1. *If $\mathbb{P}_n$ satisfies* **A2**, *then for all $n \in \mathbb{N}$, we have*

$$\mathbb{P}_n(\theta) = \int \frac{(\theta t)^n e^{-\theta t}}{n!} dF(t) \tag{25}$$

for some distribution with cdf F and $\mathbb{E}_F(x) = 1$. That is, $\mathbb{P}_n$ has a representation as a mixed Poisson distribution.

Given that any mixed Poisson distribution is invariant, we conclude that $\mathbb{P}_n$ is invariant if and only if $\mathbb{P}_n$ is a mixed Poisson distribution. This equivalence will prove to be very useful throughout the remainder of the paper. We will often switch between describing a search technology in terms of the function $\mathbb{P}_0$, and describing it in terms of the mixing distribution F.

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, the mixing distribution F is the degenerate distribution with support $\{1\}$.

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, the mixing distribution is the exponential distribution with support $[0, \infty)$ and cdf $F(x) = 1 - e^{-x}$.

4.2 Asymptotic behavior

In this paper, we are interested in what happens when θ becomes very large. Mixed Poisson distributions are especially tractable because, roughly speaking, as θ goes to infinity, the random variable $N(\theta)$ converges to the random Poisson parameter θX. We state this formally in the following lemma.

Lemma 4. *Suppose $N(\theta)$ is a random variable with an invariant distribution $\mathbb{P}_n$. As θ goes to infinity, $N(\theta)/\theta$ converges in distribution to a random variable X with distribution F, the mixing distribution of the mixed Poisson representation of $\mathbb{P}_n$.*

Lemma 4 is useful because it enables us to deduce the asymptotic properties of the random variable $N(\theta)$ directly from the properties of the mixing distribution F. For instance, the moments of $N(\theta)/\theta$ are asymptotically equal to those of F :

$$\lim_{\theta \to \infty} \mathbb{E}\left((N(\theta)/\theta)^k \right) = \mathbb{E}_F\left(X^k \right). \tag{26}$$

Equivalently, the moments of $N(\theta)$ behave asymptotically as follows:

$$\mathbb{E}_{\mathbb{P}}\left(N(\theta)^k\right) \sim_{\theta\to\infty} \theta^k \mathbb{E}_F\left(X^k\right). \tag{27}$$

This provides us with a way to calculate the *asymptotic dispersion* of $\mathbb{P}_n$. As we will see, the concept of asymptotic dispersion will turn out to be crucial for the behaviour of extreme value outcomes in our setting.

Asymptotic dispersion

To discuss asymptotic dispersion, we first need to define a measure of dispersion. It will later prove useful to define a *class* of measures of dispersion indexed by $k \in \mathbb{R}$.

First, recalling that F has mean zero, we can define a class of measures of dispersion of the mixing distribution F for any $k \in \mathbb{R}$ by

$$d_F(k) \equiv \left|\mathbb{E}_F\left(X^k\right) - 1\right|. \tag{28}$$

Analogously, we can define a class of measures of dispersion of the search technology $\mathbb{P}_n$, which has mean θ, for any $k \in \mathbb{R}$ by

$$d_{\mathbb{P}}(k;\theta) \equiv \left|\frac{\mathbb{E}_{\mathbb{P}}\left(N(\theta)^k\right) - \mathbb{E}\left(N(\theta)\right)^k}{\mathbb{E}_{\mathbb{P}}\left(N(\theta)\right)^k}\right|. \tag{29}$$

In general, for any $k \in \mathbb{R}$, the values $d_F(k)$ and $d_{\mathbb{P}}(k;\theta)$ are measures of dispersion. The reason why we need to take the absolute value in (28) and (29) is because $k \in \mathbb{R}$ and, by Jensen's inequality, we have $E_F(X^k) > 1$ if x^k is strictly convex (i.e. $k > 1$ or $k < 0$) and $E_F(X^k) < 1$ if x^k is strictly concave (i.e. $k \in (0,1)$).

In the special case where $k = 2$, we have $d_F(2) = cv_F^2$, the squared coefficient of variation of F, while $d_{\mathbb{P}}(2;\theta) = cv_{\mathbb{P}}^2$, the squared coefficient of variation of $\mathbb{P}_n$. It can be shown that $cv_{\mathbb{P}}^2 = cv_F^2 + \frac{1}{\theta}$. As the expected number of draws θ increases, the measure of dispersion $cv_{\mathbb{P}}^2$ decreases, but it may or may not converge to zero.

Given we are interested in what happens when the number of draws is large, our main object of interest will be the *asymptotic dispersion*, which is defined as the "residual" dispersion that remains in the limit as $\theta \to \infty$. Defining the *asymptotic dispersion* of $\mathbb{P}_n$ by $d_{\mathbb{P}}(k) \equiv \lim_{\theta\to\infty} d_{\mathbb{P}}(k;\theta)$, we can now state the following result.

Corollary 2. *For any $k \in \mathbb{R}$, the asymptotic dispersion $d_{\mathbb{P}}(k)$ of the search technology $\mathbb{P}_n$ is equal to the dispersion $d_F(k)$ of the mixing distribution F. That is,*

$$d_{\mathbb{P}}(k) = d_F(k). \tag{30}$$

The asymptotic dispersion $d_{\mathbb{P}}(k)$ is zero for all $k \in \mathbb{R}$ if and only if $\mathbb{P}_n$ is Poisson.

Corollary 2 tells us that $d_F(k)$ is *both* a measure of dispersion for the mixing distribution F and a measure of *asymptotic dispersion* $d_{\mathbb{P}}(k)$ for the search technology $\mathbb{P}_n$. This will prove useful later as the asymptotic dispersion will turn out to be what matters for the impact of the search technology on extreme value outcomes.

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, then F is degenerate and $d_F(2) = cv_F^2 = 0$, so the asymptotic dispersion is zero as $\theta \to \infty$.

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, then F is exponential and $d_F(2) = cv_F^2 = 1$, so there is asymptotic dispersion as $\theta \to \infty$, which is equal to that of the mixing distibution F.

In the special case where $\mathbb{P}_n$ is Poisson, the random variable $N(\theta)$ behaves asymptotically like a constant (i.e. as though $\mathbb{P}_n$ is asymptotically degenerate). However, if the search technology $\mathbb{P}_n$ is not Poisson, i.e. the mixing distribution F is non-degenerate, then $N(\theta)$ exhibits asymptotic dispersion. This will turn out to be crucial for understanding our next results regarding the extreme value distribution.

5 Extreme value distribution

In this section, we prove our first main theorem. We provide a general result regarding the extreme value distribution when the number of draws from the underlying distribution is random and the search technology is invariant.

We assume throughout the remainder of the paper that G is well-behaved and has a tail index $\gamma \in \mathbb{R}$. The tail index is a measure of tail fatness, with a higher value of γ corresponding to fatter tails. The tail index is crucial for determining the domain of attraction of G, i.e. the type of extreme value distribution. It is defined below.

Definition 2. *We say that G is* well-behaved *if and only if* $\lim_{x\to\overline{x}} \frac{1-G(x)}{g(x)} = a$ *where* $a \in \mathbb{R}^+ \cup \{+\infty\}$ *and G has finite* tail index $\gamma \in \mathbb{R}$ *given by*

$$\lim_{x\to\overline{x}} \frac{d}{dx}\left(\frac{1-G(x)}{g(x)}\right) = \gamma. \tag{31}$$

Let $X_1, ..., X_n$ be i.i.d. random variables with distribution G. Define the random variable $M_n \equiv \max\{X_1, ..., X_n\}$ where $n \geq 1$. Classical extreme value theory tells us that under certain conditions on G, there exist normalizing constants a_n, b_n such that the sequence of normalized random variables $Z_n = a_n M_n + b_n$ converges in distribution as $n \to \infty$. Moreover, the Fisher-Tippett-Gnedenko extreme value theorem tells us that the limiting distribution of Z_n must take the following form:

$$H_\gamma(x) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}} & \gamma \neq 0 \\ e^{-e^{-x}} & \gamma = 0 \end{cases} \tag{32}$$

where $H_\gamma(x) \equiv \lim_{n\to\infty} \Pr(Z_n \leq x)$. This means that the extreme value distribution must be one of only three types: it must be either Frechet (if $\gamma > 0$), Weibull (if $\gamma < 0$), or Gumbel (if $\gamma = 0$). Henceforth, we define $v_\gamma(x) \equiv (1+\gamma x)^{-1/\gamma}$ if $\gamma \neq 0$ and $v_\gamma(x) \equiv e^{-x}$ if $\gamma = 0$. See, for example, de Haan and Ferreira (2007) for the relevant conditions and normalizing constants for different distributions G.

Now suppose $N(\theta)$ is a discrete random variable with mean θ, and define the random variable $M_{N(\theta)} \equiv \max\{X_1, ..., X_{N(\theta)}\}$. The following theorem tells us that there exist normalizing constants a_θ, b_θ such that the sequence of normalized random variables $Z_{N(\theta)} = a_\theta M_{N(\theta)} + b_\theta$ converges in distribution as $\theta \to \infty$. However, Theorem 1 says that the limiting distribution takes a form which depends not only on the underlying distribution G and its tail index, but also on the *search technology* $\mathbb{P}_n$.

Theorem 1. *Suppose that $\mathbb{P}_n$ satisfies* **A2**. *Suppose also there exist normalizing constants a_n, b_n such that the distribution of the normalized random variable $Z_n = a_n M_n + b_n$ converges as $n \to \infty$ to the extreme value distribution*

$$H_\gamma(x) = e^{-v_\gamma(x)} \tag{33}$$

where $H_\gamma(x) \equiv \lim_{n\to\infty} \Pr(Z_n \leq x)$. If $N(\theta)$ is a random variable with distribution $\mathbb{P}_n$ and mean θ, then the distribution of the normalized random variable $Z_{N(\theta)} = a_\theta M_{N(\theta)} +$

b_θ converges as $\theta \to \infty$ to the following extreme value distribution:

$$H_{\gamma,\mathbb{P}}(x) = \mathbb{P}_0(\upsilon_\gamma(x)) \tag{34}$$

where $H_{\gamma,\mathbb{P}}(x) \equiv \lim_{\theta\to\infty} \Pr(Z_{N(\theta)} \leq x)$.

In the special case where $\mathbb{P}_n$ is Poisson and $\mathbb{P}_0(x) = e^{-x}$, it is easy to see that we recover the original extreme value distribution, i.e. $H_{\gamma,\mathbb{P}}(x) = H_\gamma(x)$. Conversely, if $H_{\gamma,\mathbb{P}}(x) = H_\gamma(x)$ then $\mathbb{P}_n$ must be Poisson (if it satisfies **A2**).

Corollary 3. *Suppose that $\mathbb{P}_n$ satisfies* **A2**. *The extreme value distribution is*

$$H_{\gamma,\mathbb{P}}(x) = H_\gamma(x) \tag{35}$$

if and only if the search technology $\mathbb{P}_n$ is Poisson.

This result explains why the Poisson distribution has a special role in extreme value theory. When we move away from the Poisson distribution, however, the extreme value distribution no longer takes any of the standard forms.

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, the extreme value distribution is

$$H_{\gamma,\mathbb{P}}(x) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}} & \gamma \neq 0 \\ e^{-e^{-x}} & \gamma = 0 \end{cases} \tag{36}$$

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

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, the extreme value distribution is

$$H_{\gamma,\mathbb{P}}(x) = \frac{1}{1 + \upsilon_\gamma(x)}. \tag{37}$$

where $\upsilon_\gamma(x) \equiv (1+\gamma x)^{-1/\gamma}$ if $\gamma \neq 0$ and e^{-x} if $\gamma = 0$. This differs from any of the three standard forms of the extreme value distribution (Fréchet, Gumbel, Weibull).

6 Extreme value outcomes

In this section, we prove our second main theorem. We provide a general result regarding the outcomes of extreme value processes, e.g. functions of maxima, when the number of draws is random and the search technology is invariant. This result generalizes a result in Gabaix et al. (2016) regarding extreme value outcomes.

More precisely, we study the asymptotic behaviour of the expected value of a function ζ of the maximum. Specifically, we take the expectation of $\zeta(x)$ with regard to the distribution of the maximum, $H_{\mathbb{P}}(x;\theta)$, and we then consider the limit as θ goes to infinity. That is, we consider the asymptotic behavior of the function $\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x)$. We refer to the asymptotic behavior of $\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x))$ as an *extreme value outcome*.

Before presenting our main result, we require the following definition.[5]

Definition 3. *A function $h : \mathbb{R}_+ \to \mathbb{R}$ is* regularly varying at zero with index ρ, *denoted $h \in RV_\rho^0$, if and only if, for all $\lambda > 0$, we have*

$$\lim_{t\to 0} \frac{h(\lambda t)}{h(t)} = \lambda^\rho. \tag{38}$$

The following assumption regarding the function ζ will be used in Theorem 2.

Assumption 3 (**A3**). *The function $\zeta : [\underline{x}, \bar{x}] \to \mathbb{R}$ is measurable and bounded on $[\underline{x}, x]$ for any $x \in (\underline{x}, \bar{x})$, and the integral $\int_{\underline{x}}^{\overline{x}} |\zeta(x)g(x)|\, dx$ is finite.*

We also define the Mellin transform of a function, which will prove useful. The *Mellin transform* of a function $k : \mathbb{R}^+ \to \mathbb{R}$, denoted $\breve{k}$, is defined by

$$\breve{k}(a) := \int_0^\infty t^{a-1} k(t) dt. \tag{39}$$

In the statement of Theorem 2 and the following results, we adopt standard notation and write $F_1(\theta) \sim_{\theta\to\infty} F_2(\theta)$, or simply $F_1(\theta) \sim F_2(\theta)$, if and only if $\lim_{\theta\to\infty} \frac{F_1(\theta)}{F_2(\theta)} = 1$.

Theorem 2. *Suppose that $\mathbb{P}_n$ satisfies* **A2** *and ζ satisfies* **A3** *and $\zeta(G^{-1}(1-t)) \in RV_\rho^0$ for some $\rho > -1$. If $\mathbb{E}_F(X^a)$ is finite for all a in a neighbourhood of $-\rho$, then*

$$\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x)) = \int_{\underline{x}}^{\overline{x}} \zeta(x) h_{\mathbb{P}}(x;\theta) dx \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right) \Psi_{\mathbb{P}}(\rho+1) \tag{40}$$

[5] See Bingham, Goldie, and Teugels (1987) or Resnick (1987).

where $\Psi_{\mathbb{P}} : \mathbb{R} \to \mathbb{R}^+$ *is defined by*

$$\Psi_{\mathbb{P}}(a) := \int_0^\infty z^{a-1}(-\mathbb{P}_0'(z))dz. \tag{41}$$

The formal proof of Theorem 2 can be found in the Appendix, but we include a simple heuristic derivation below, which highlights the key role of regular variation.

Heuristic proof. By the assumption of invariance, the pdf of $H_{\mathbb{P}}(x;\theta)$ is

$$h_{\mathbb{P}}(x;\theta) = \frac{-g(x)\theta\mathbb{P}_0'(\theta(1-G(x)))}{1-\mathbb{P}_0(\theta)} \tag{42}$$

Given that $\lim_{\theta\to\infty}\mathbb{P}_0(\theta) = 0$, and changing variables by setting $z = \theta(1-G(x))$,

$$\int_{\underline{x}}^{\overline{x}} \zeta(x)h_{\mathbb{P}}(x;\theta)dx \sim_{\theta\to\infty} \int_0^\theta \zeta\left(G^{-1}\left(1-\frac{z}{\theta}\right)\right)(-\mathbb{P}_0'(z))dz. \tag{43}$$

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

$$\lim_{t\to 0}\frac{\zeta\left(G^{-1}\left(1-zt\right)\right)}{\zeta\left(G^{-1}\left(1-t\right)\right)} = z^\rho. \tag{44}$$

Setting $t = 1/\theta$, this is equivalent to

$$\zeta\left(G^{-1}\left(1-\frac{z}{\theta}\right)\right) \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right) z^\rho. \tag{45}$$

This allows us to take the first term on the right outside the integral:

$$\int_0^\theta \zeta\left(G^{-1}\left(1-\frac{z}{\theta}\right)\right)(-\mathbb{P}_0'(z))dz \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)\int_0^\infty z^\rho(-\mathbb{P}_0'(z))dz. \tag{46}$$

Theorem 2 shows clearly that extreme value outcomes may be affected by the search technology $\mathbb{P}_n$. To understand this better, the following result provides a useful expression for the term $\Psi_{\mathbb{P}}(a)$ that will help us understand exactly *how* the search technology influences extreme value outcomes.

Lemma 5. *If* $\mathbb{P}_n$ *satisfies* **A2**, *then*

$$\Psi_{\mathbb{P}}(a) = \Gamma(a)\varphi_{\mathbb{P}}(1-a) \tag{47}$$

where $\varphi_{\mathbb{P}}: \mathbb{R} \to \mathbb{R}^+$ *is defined for all* $k \in \mathbb{R}$ *by*

$$\varphi_{\mathbb{P}}(k) \equiv \mathbb{E}_F(X^k) \tag{48}$$

and $\Gamma(.)$ *is the Gamma function, defined by* $\Gamma(a) \equiv \int_0^\infty t^{a-1}e^{-t}dt$.

The corollary below follows immediately from Theorem 2 and Lemma 5.

Corollary 4. *Suppose that* $\mathbb{P}_n$ *satisfies* **A2** *and* ζ *satisfies* **A3** *and* $\zeta(G^{-1}(1-t)) \in RV_\rho^0$ *for some* $\rho > -1$. *If* $\mathbb{E}_F(X^a)$ *is finite for all* a *in a neighbourhood of* $-\rho$, *then*

$$\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x)) \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)\Gamma(\rho+1) \underbrace{\varphi_{\mathbb{P}}(-\rho)}_{\text{effect of search frictions}} \tag{49}$$

where $\Gamma(a) \equiv \int_0^\infty t^{a-1}e^{-t}dt$ *and* $\varphi_{\mathbb{P}}(k) \equiv \mathbb{E}_F(X^k)$.

Corollary 4 makes it easy to compare our general result with the result proven in Gabaix et al. (2016). If the distribution $\mathbb{P}_n$ is degenerate, they show that

$$\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x)) \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)\Gamma(\rho+1). \tag{50}$$

This corresponds to the special case of (49) where there is no effect of search frictions.

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, then F is degenerate and $\mathbb{E}_F(X^{-\rho}) = 1$, so $\varphi_{\mathbb{P}}(-\rho) = 1$ by Lemma 5 and thus

$$\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x)) \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)\Gamma(\rho+1) \tag{51}$$

by Corollary 4. We therefore recover the result in Mangin (2022), which holds for the special case where the search technology $\mathbb{P}_n$ is Poisson with mean θ.

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, then F is exponential and $\mathbb{E}(X^{-\rho}) = \Gamma(1-\rho)$, so $\varphi_{\mathbb{P}}(-\rho) = \Gamma(1-\rho)$ by Lemma 5 and thus we obtain

$$\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x)) \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)\Gamma(\rho+1) \underbrace{\Gamma(1-\rho)}_{\text{effect of search frictions}} \tag{52}$$

by Corollary 4. There is clearly an asymptotic effect of search frictions on extreme value outcomes whenever ρ is non-zero, but the direction of the effect depends on the value of ρ, which we discuss in more detail next.

Effect of asymptotic dispersion

There is a very close connection between $\varphi_{\mathbb{P}}(k) \equiv \mathbb{E}_F(X^k)$ and the measure of dispersion we defined earlier, $d_F(k) = \left|\mathbb{E}_F\left(X^k\right) - 1\right|$. Using the fact that $\varphi_{\mathbb{P}}(k) \equiv \mathbb{E}_F(X^k)$, we can write $d_F(k) = |\varphi_{\mathbb{P}}(k) - 1|$. Recall that, by Jensen's inequality, we have $E_F(X^k) > 1$ if X^k is strictly convex (i.e. $k > 1$ or $k < 0$) and $E_F(X^k) < 1$ if X^k is strictly concave (i.e. $k \in (0,1)$). Since we assume that $\rho > -1$, we can express the effect of search frictions $\varphi_{\mathbb{P}}(-\rho)$ as follows:

$$\varphi_{\mathbb{P}}(-\rho) = \begin{cases} 1 + d_F(-\rho)), & \rho \geq 0, \\ 1 - d_F(-\rho), & \rho < 0. \end{cases} \tag{53}$$

This makes it clear that extreme value outcomes may be either increasing or decreasing in the dispersion of the mixing distribution F, depending on both the specific application and the underlying distribution G, which will jointly determine the relevant value of ρ. We know from Corollary 2 that $d_F(k)$ is also equal to the *asymptotic dispersion* $d_{\mathbb{P}}(k)$ of the distribution $\mathbb{P}_n$. That is, $\lim_{\theta\to\infty} d_{\mathbb{P}}(k;\theta) = d_F(k)$. So, it is equivalent to say that extreme value outcomes may be either increasing or decreasing in the asymptotic dispersion of the distribution $\mathbb{P}_n$.

The following corollary summarizes the asymptotic effects of the search technology on extreme value outcomes.

Corollary 5. *Suppose that $\mathbb{P}_n$ satisfies* **A2** *and ζ satisfies* **A3** *and $\zeta(G^{-1}(1-t)) \in RV_\rho^0$ for some $\rho > -1$. If $\mathbb{E}_F(X^a)$ is finite for all a in a neighbourhood of $-\rho$, then*

1. *If $\rho = 0$, there is no asymptotic effect of search frictions on extreme value outcomes regardless of the search technology $\mathbb{P}_n$.*

2. *If $\rho \neq 0$, there is an asymptotic effect of search frictions on extreme value outcomes if and only if there is asymptotic dispersion, $d_{\mathbb{P}}(-\rho) > 0$.*

3. *If $\rho > 0$, extreme value outcomes are increasing in asymptotic dispersion.*

4. If $\rho < 0$, extreme value outcomes are decreasing in asymptotic dispersion.

In the next section, we turn to considering some applications of our general results.

7 Applications

In this section, we discuss some applications of our general results. We consider two applications: aggregate productivity and markups. First, we discuss a particular family of search technologies that will be useful for our applications.

7.1 Negative binomial family

For both of our applications, we focus on the negative binomial family of search technologies. All of the distributions in this family satisfy Assumption **A2**.

This family of search technologies is very useful because, as we will see, it nests the Poisson search technology as a limiting case, and it nests the Geometric search technology as a special case. Moreover, the family is analytically tractable and may be useful for empirical applications because it features a single parameter, r, that captures the degree of asymptotic dispersion and could potentially be estimated.

We can interpret the negative binomial random variable $N_r(\theta)$ as counting the number n of failures before r successes, where the probability of success is $\frac{r}{r+\theta}$ and the expected number of failures before r successes is θ. Alternatively, it can be interpreted as a mixed Poisson distribution, with F equal to the Gamma distribution.

If $\mathbb{P}_n(\theta)$ is negative binomial, then

$$\mathbb{P}_n(\theta) = \binom{n+r-1}{n} \left(\frac{r}{r+\theta}\right)^r \left(\frac{\theta}{r+\theta}\right)^n \tag{54}$$

where $r \in \mathbb{N} \setminus \{0\}$ and the function $\mathbb{P}_0$ is given by

$$\mathbb{P}_0(z) = \left(\frac{r}{r+z}\right)^r. \tag{55}$$

If $\mathbb{P}_n$ is in the negative binomial family, the degree of asymptotic dispersion $d_{\mathbb{P}}(k)$ is captured by the parameter r. In particular, if $k = 2$, the asymptotic dispersion is

$$cv_\infty^2 = \frac{1}{r}. \tag{56}$$

Equivalently, $1/r$ is also equal to cv_F^2, the squared coefficient of variation of the mixing distribution F. This means that we can think of $1/r$ as representing the "residual" degree of search frictions when the expected number of firms becomes large. This can vary between zero and one. By varying the single parameter r of the negative binomial family of distributions, we can determine the effect the search technology.

Observe that when $r = 1$, we obtain the geometric distribution. In the limit as $r \to \infty$, we obtain the Poisson distribution, which has zero asymptotic dispersion.

7.2 Aggregate productivity

Suppose there is a continuum of measure one of firms. Firm-level productivity is a random shock that is equal to the firm's highest draw from an underlying distribution of productivities G that has unbounded upper support. Each firm has the same expected number of productivity draws θ, but the actual number of draws each firm gets is random and given by a search technology $\mathbb{P}_n$ that satisfies **A2**.

Aggregate productivity $y_{\mathbb{P}}(\theta)$ is equal to the average firm-level productivity. That is, it is equal to the expected value of each firm's highest draw from the underlying productivity distribution G, where the number of draws $n \sim \mathbb{P}_n(\theta)$, i.e. $y_{\mathbb{P}}(\theta) \equiv \mathbb{E}_{H_{\mathbb{P}}}(x)$. For any $\theta \in \Theta$, aggregate productivity is given by

$$y_{\mathbb{P}}(\theta) = \int_{\underline{x}}^{\infty} x \, dH_{\mathbb{P}}(x). \tag{57}$$

Now suppose the expected number of productivity draws for each firm becomes large. The next proposition provides an asymptotic result for aggregate productivity in the limit as $\theta \to \infty$. This result follows directly from Theorem 2 with $\zeta(x) = x$, which implies $\rho = -\gamma$, where γ is the tail index of G.

Proposition 3. *If $\mathbb{P}_n$ satisfies* **A2**, *G has tail index $\gamma < 1$ and $\bar{x} = \infty$, and $\mathbb{E}_F(X^a)$ is finite for all a in a neighbourhood of γ, then then*

1. *Aggregate productivity is given by*

$$y_{\mathbb{P}}(\theta) \sim_{\theta \to \infty} G^{-1}\left(1 - \frac{1}{\theta}\right) \Gamma(1-\gamma) \qquad \underbrace{\varphi_{\mathbb{P}}(\gamma).}_{\text{effect of search frictions}} \tag{58}$$

2. *For all $\gamma \in (0,1)$, $y_{\mathbb{P}}(\theta)$ is decreasing in asymptotic dispersion $d_F(\gamma)$.*

3. If $\gamma = 0$, then $y_{\mathbb{P}}(\theta)$ does not depend on the asymptotic dispersion $d_F(\gamma)$.

Proposition 3 implies that, for any given expected number of draws θ, aggregate productivity is maximized when the asymptotic dispersion is zero, i.e. the search technology is Poisson. If the search technology is more dispersed than Poisson, this implies a kind of "misallocation" in terms of the distribution of the number of productivity draws across firms, which reduces aggregate productivity. Surprisingly, this misallocation due to search frictions still matters even in the limit as the expected number of productivity draws each firm gets becomes large.

We now turn to the negative binomial family of search technologies.

Corollary 6. *If $\mathbb{P}_n$ is negative binomial with $r > -\gamma$, where $\gamma < 1$ and $\bar{x} = \infty$, then*

1. *Aggregate productivity is given by*

$$y_{\mathbb{P}}(\theta) \sim_{\theta \to \infty} G^{-1}\left(1 - \frac{1}{\theta}\right) \Gamma(1-\gamma) \underbrace{\left(\frac{r^{-\gamma}\Gamma(r+\gamma)}{\Gamma(r)}\right)}_{\text{effect of search frictions}}. \tag{59}$$

2. *For all $\gamma \in (0, 1)$, $y_{\mathbb{P}}(\theta)$ is decreasing in $1/r$.*

3. *If $\gamma = 0$, then $y_{\mathbb{P}}(\theta)$ does not depend on $1/r$.*

Importantly, search frictions can affect not only the *level* of aggregate productivity but also the cross-sectional *distribution* of productivity across firms. For the negative binomial family of search technologies, the cross-sectional distribution of productivity across firms is given by the following result. To obtain this result, we directly apply Theorem 1, which says that $H_{\gamma,\mathbb{P}}(x) = \mathbb{P}_0(v_\gamma(x))$.

Corollary 7. *If $\mathbb{P}_n$ is negative binomial with parameter $r > -\gamma$, the cdf of the (normalized) cross-sectional productivity distribution is*

$$H_{\gamma,\mathbb{P}}(x) = \left(\frac{r}{r + v_\gamma(x)}\right)^r \tag{60}$$

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

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, letting $r \to \infty$ yields

$$y_{\mathbb{P}}(\theta) \sim_{\theta \to \infty} G^{-1}\left(1 - \frac{1}{\theta}\right)\Gamma(1-\gamma). \tag{61}$$

That is, the Poisson search technology yields the same aggregate productivity as a degenerate search technology, where each firm has the same number of draws.

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

$$H_{\gamma,\mathbb{P}}(x) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}} & \gamma \neq 0 \\ e^{-e^{-x}} & \gamma = 0 \end{cases} \tag{62}$$

which is just the standard form of the extreme value distribution.

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, setting $r = 1$ delivers

$$y_{\mathbb{P}}(\theta) \sim_{\theta \to \infty} G^{-1}\left(1 - \frac{1}{\theta}\right)\Gamma(1-\gamma)\Gamma(1+\gamma). \tag{63}$$

In this case, the search technology has no effect if the tail index is $\gamma = 0$, but it leads to lower aggregate productivity if $\gamma > 0$ because $\Gamma(1+\gamma) < 1$. This effect occurs despite the fact that the expected number of productivity draws each firm receives is the *same* value θ, and we are also taking the limit as θ becomes large.

The cross-sectional productivity distribution is also affected by the search technology. The cdf of the (normalized) cross-sectional productivity distribution is

$$H_{\gamma,\mathbb{P}}(x) = \frac{1}{1 + v_\gamma(x)} \tag{64}$$

which is different from all three standard extreme value distributions.

Example: Pareto distribution of productivities

To work out the exact expressions for aggregate productivity and the cross-sectional productivity distribution across all firms, we need to consider a specific underlying distribution from which the firm-level productivities are drawn.

Suppose the underlying distribution of productivities is Pareto, $G(x) = 1 - x^{-1/\gamma}$ where $\gamma \in (0, 1)$. If all firms have the same number of draws θ, i.e. if the search technology $\mathbb{P}_n$ is degenerate, we know from Section 3 that the (normalized) cross-sectional distribution of productivities is Fréchet. More precisely, $H_\gamma(x) = e^{-x^{-1/\gamma}}$. Next, using the fact that $G^{-1}\left(1 - \frac{1}{\theta}\right) = \theta^\gamma$, aggregate productivity is given by

$$y_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \theta^\gamma \Gamma(1-\gamma). \tag{65}$$

Now suppose the number of draws is given by a search technology $\mathbb{P}_n$ that is negative binomial with parameter $r > -\gamma$. Aggregate productivity is given by

$$y_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \theta^\gamma \Gamma(1-\gamma) \underbrace{\left(\frac{r^{-\gamma}\Gamma(r+\gamma)}{\Gamma(r)}\right)}_{\textit{effect of search frictions}} \tag{66}$$

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

$$H_{\gamma,\mathbb{P}}(x) = \left(\frac{r}{r + x^{-1/\gamma}}\right)^r. \tag{67}$$

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, letting $r \to \infty$ yields

$$y_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \theta^\gamma \Gamma(1-\gamma) \tag{68}$$

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

$$H_{\gamma,\mathbb{P}}(x) = e^{-x^{-1/\gamma}}. \tag{69}$$

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, setting $r = 1$ delivers

$$y_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \theta^\gamma \Gamma(1-\gamma)\Gamma(1+\gamma) \tag{70}$$

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

$$H_{\gamma,\mathbb{P}}(x) = \frac{1}{1 + x^{-1/\gamma}}. \tag{71}$$

From Corollary 6, we know that aggregate productivity $y_{\mathbb{P}}(\theta)$ is strictly decreasing in the asymptotic dispersion $1/r$ because $\gamma \in (0, 1)$. Therefore, aggregate productivity

$y_{\mathbb{P}}(\theta)$ is highest when $r \to \infty$ (Poisson) and lowest when $r = 1$ (Geometric). From the examples above, the ratio of the lowest to highest aggregate productivity is $\Gamma(1+\gamma)$.

For example, suppose that $\gamma = 0.5$. If all firms have the same number of productivity draws, or alternatively if the search technology $\mathbb{P}_n$ is Poisson, then $y_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \theta^\gamma\sqrt{\pi}$. However, if $\mathbb{P}_n$ is a Geometric search technology then $y_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \theta^\gamma(\pi/2)$. So, the ratio of aggregate productivity for Geometric versus Poisson is $\sqrt{\pi}/2 \approx 0.886$. This means that the decline in aggregate productivity due to the frictional search technology is significant: around 11.4%. This is despite the fact that all firms have the same expected number of productivity draws and we are taking the limit as the number of productivity draws per firm becomes large.

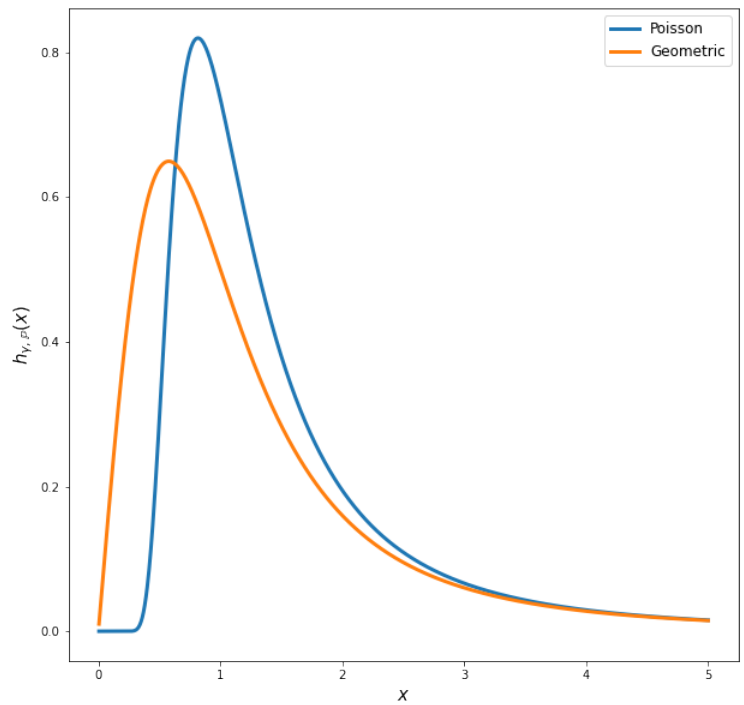

Figure 1: Cross-sectional firm productivity distribution when $\gamma = 0.5$.

There is also a significant effect of search frictions on cross-sectional productivity dispersion. Consider the measure of cross-sectional productivity dispersion $cv_{H,\mathbb{P}}$, defined as the coefficient of variation of the distribution $H_{\gamma,\mathbb{P}}$. Suppose that $\gamma = 0.5$. If all firms have the same number of productivity draws, or the search technology $\mathbb{P}_n$ is Poisson, then $cv_{H,\mathbb{P}} = 25.03$. However, if $\mathbb{P}_n$ is a Geometric search technology then $cv_{H,\mathbb{P}} = 32.82$. Thus, the increase in cross-sectional productivity dispersion due to the frictional search technology is significant: around 28%.

Figure 1 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. Figure 1 shows that when the search technology changes from Poisson to Geometric, average productivity decreases and cross-sectional productivity dispersion increases.

7.3 Markups

Consider a discrete choice model with random utility shocks. There is a continuum of measure one of consumers, and a continuum of measure θ of firms. Each firm sells a single indivisible good, and each consumer has unit demand.

Each consumer meets a random number n of firms. The randomness of n reflects the presence of search frictions. We assume that the distribution of n across consumers as given by a search technology $\mathbb{P}_n$ with mean θ that satisfies **A2**.

Each consumer draws a random *utility shock* x_i from a distribution G for each firm i they meet. This shock x_i represents the consumer's valuation of firm i's good. These utility shocks are observed by the consumer and each of the n firms a consumer meets.

After observing the utility shocks $(x_1, ..., x_n)$, as well as the number of competitors, firms set prices simultaneously. Let M_n denote the maximum of $(x_1, ..., x_n)$ and let S_n denote the second highest utility shock. In equilibrium, each consumer purchases from the firm which gives it the highest utility, and the markup, i.e. price minus marginal cost, is equal to $\mu_n = M_n - S_n$. This type of pricing is sometimes called either "limit pricing" (or personalized pricing) or Bertrand competition.

The average markup across consumers (given $n \geq 1$) is

$$\mu_{\mathbb{P}}(\theta) = \frac{\sum_{n=1}^{\infty} \mathbb{P}_n(\theta)\mu_n}{1 - \mathbb{P}_0(\theta)} \tag{72}$$

which can be expressed as follows:

$$\mu_{\mathbb{P}}(\theta) = \int_{\underline{x}}^{\overline{x}} \left(\frac{1 - G(x)}{g(x)} \right) h_{\mathbb{P}}(x; \theta) dx. \tag{73}$$

We are interested in the asymptotic markup in the limit as the expected number of firms each consumer meets becomes large, i.e. as $\theta \to \infty$, which is

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \mathbb{E}_{H_{\mathbb{P}}}\left(\frac{1-G(x)}{g(x)}\right). \tag{74}$$

To prove the following result, we apply Theorem 2 with $\zeta(x) = \frac{1-G(x)}{g(x)}$.

Proposition 4. *If $\mathbb{P}_n$ satisfies* **A2**, *and G has tail index $\gamma < 1$, then*

1. *The asymptotic markup is*

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{\Gamma(1-\gamma)}{\theta g\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)} \underbrace{\varphi_{\mathbb{P}}(\gamma).}_{\text{effect of search frictions}} \tag{75}$$

2. *In the limit as $\theta \to \infty$, we have $\mu_{\mathbb{P}}(\theta) \to 0$ if and only if $\lim_{x\to\bar{x}} \frac{1-G(x)}{g(x)} = 0$.*

Corollary 8 below summarizes the effect of asymptotic dispersion on markups. Interestingly, the effect may be either positive or negative depending on the tail index of the underlying distribution G. That is, greater dispersion of the search technology, i.e. higher search frictions, can either increase or decrease markups.

Corollary 8. *If $\mathbb{P}_n$ satisfies* **A2**, *and G has tail index $\gamma < 1$, then*

1. *For all $\gamma \in (0, 1)$, $\mu_{\mathbb{P}}(\theta)$ is decreasing in asymptotic dispersion $d_F(\gamma)$.*
2. *For all $\gamma \in [-1, 0)$, $\mu_{\mathbb{P}}(\theta)$ is increasing in asymptotic dispersion $d_F(\gamma)$.*
3. *If $\gamma = 0$, then $\mu_{\mathbb{P}}(\theta)$ does not depend on the asymptotic dispersion $d_F(\gamma)$.*

We now consider the negative binomial family of search technologies. Corollary 9 provides an expression for the asymptotic markup and some comparative statics describing the effect of an increase in $1/r$ (i.e. an increase in cv_∞^2).

Corollary 9. *If $\mathbb{P}_n$ is negative binomial with $r > -\gamma$, and $\gamma < 1$, then*

1. *For all $\gamma \in [-1, 1)$, the asymptotic markup is*

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{\Gamma(1-\gamma)}{\theta g\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)} \underbrace{\left(\frac{r^{-\gamma}\Gamma(r+\gamma)}{\Gamma(r)}\right)}_{\text{effect of search frictions}}.$$

2. *For all* $\gamma \in (0,1)$*,* $\mu_{\mathbb{P}}(\theta)$ *is decreasing in* $1/r$*.*

3. *For all* $\gamma \in [-1,0)$*,* $\mu_{\mathbb{P}}(\theta)$ *is increasing in* $1/r$*.*

4. *If* $\gamma = 0$*, then* $\mu_{\mathbb{P}}(\theta)$ *does not depend on* $1/r$*.*

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, letting $r \to \infty$ yields

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{\Gamma(1-\gamma)}{\theta g\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)}. \tag{76}$$

In this case, we recover the same expression as in Gabaix et al. (2016). That is, the asymptotic markup is the same as when n is deterministic.

Example 2. If $\mathbb{P}_n$ is a Geometric search technology, setting $r = 1$ delivers

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{\Gamma(1-\gamma)}{\theta g\left(G^{-1}\left(1-\frac{1}{\theta}\right)\right)} \quad \underbrace{\Gamma(1+\gamma).}_{\textit{effect of search frictions}} \tag{77}$$

In this case, there is a clear effect of asymptotic search frictions on markups (unless $\gamma = 0$), but the direction of this effect depends on the tail index γ. Markups for the Geometric search technology may be either higher or lower than for the Poisson search technology depending on the value of γ.

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 meets the same number of firms θ, i.e. if the search technology $\mathbb{P}_n$ is degenerate, we know from Gabaix et al. (2016) that

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{1}{\theta}. \tag{78}$$

As expected, the average markup is strictly decreasing in the number of firms.

Now suppose the search technology $\mathbb{P}_n$ is negative binomial with parameter $r > -\gamma$.

In this case, the average markup is given by

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{1}{\theta} \underbrace{\left(\frac{r}{r-1}\right)}_{\text{effect of search frictions}}. \tag{79}$$

For any value of r, the average markup goes to zero in the limit as the expected number of firms each consumer meets becomes large. However, we can still consider the *ratio* of asymptotic markups for different search technologies. We know that the asymptotic markup is increasing in the asymptotic dispersion $1/r$ by Corollary 9 because $\gamma = -1$ in this example. Therefore, the asymptotic markup $\mu_{\mathbb{P}}(\theta)$ is highest when $r \to \infty$ (Poisson) and lowest as $r \to 1$.

Example 1. If $\mathbb{P}_n$ is a Poisson search technology, letting $r \to \infty$ yields

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{1}{\theta} \tag{80}$$

which is idential to the asymptotic markup for the case where $\mathbb{P}_n$ is degenerate.

Example 2. If $\mathbb{P}_n$ is negative binomial with $r = 2$, we obtain

$$\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} \frac{2}{\theta}. \tag{81}$$

Note that we cannot consider the Geometric example because we require $r > -\gamma$ and $\gamma = -1$ if the underlying distribution G is uniform.

If all consumers meet the same number of firms (i.e. $\mathbb{P}_n$ is degenerate), or the search technology $\mathbb{P}_n$ is Poisson, the asymptotic markup is $\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} 1/\theta$. However, if $\mathbb{P}_n$ is negative binomial with $r = 2$ then $\mu_{\mathbb{P}}(\theta) \sim_{\theta\to\infty} 2/\theta$, which is twice as high. Therefore, the increase in markups due to the frictional search technology is 100%. This is despite the fact that the expected number of firms each consumer meets is the same in both cases, and we are taking the limit as the expected number of firms becomes large.

8 Conclusion

This paper provides some general results regarding the asymptotic effect of search frictions on the outcomes of extreme value processes, which are widespread in economics. To do this, we allow the number of draws from the *underlying distribution* (e.g. of productivities, efficiencies, or ideas) to be given by a discrete probability distribution called the *search technology*. We show that extreme value outcomes, and the nature of the extreme value distribution itself, depend not only on the underlying distribution and its tail index, but also on properties of the search technology.

For example, the fact that the Pareto distribution gives rise to the Fréchet extreme value distribution is widely used in economic applications. For the class of search technologies we consider, we show that the Pareto distribution gives rise to an extreme value distribution that is Fréchet if and only if the search technology is either Poisson or degenerate. In general, for any other search technologies in the class we consider, the extreme value distribution does not take any of the three standard types given by classical extreme value theory (Fréchet, Gumbel, or Weibull).

We find that extreme value outcomes may be either increasing or decreasing in the asymptotic dispersion of the search technology, depending on both the tail index of the underlying distribution and the specific application. We consider some applications of our results to both aggregate productivity and markups. We find that the nature of the search technology can have a quantitatively significant effect on both aggregate productivity and the degree of cross-sectional productivity dispersion. We also find that the search technology can significantly affect asymptotic markups.

References

O. Barndorff-Nielsen. On the limit distribution of the maximum of a random number of independent random variables. *Acta Math. Acad. Sci. Hungar*, 15:399–403, 1964.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

J. Boehm and E. Oberfield. Misallocation in the Market for Inputs: Enforcement and the Organization of Production. *The Quarterly Journal of Economics*, 135(4): 2007–2058, 06 2020.

J. Boehm and E. Oberfield. Growth and the Fragmentation of Production. *Working paper*, 2022.

F. J. Buera and R. E. Lucas. Idea flows and economic growth. *Annual Review of Economics*, 10(1):315–345, 2018.

F. J. Buera and E. Oberfield. The global diffusion of ideas. *Econometrica*, 88(1): 83–114, 2020.

X. Cai, P. Gautier, and R. Wolthoff. Note on invariant distributions. *Working paper*, 2022.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2007.

J. Eaton and S. Kortum. International technology diffusion: Theory and measurement. *International Economic Review*, 40(3):537–570, 1999.

X. Gabaix, D. Laibson, D. Li, H. Li, S. Resnick, and C. G. de Vries. The impact of competition on prices with numerous firms. *Journal of Economic Theory*, 165(C): 1–24, 2016.

C. I. Jones. The Shape of Production Functions and the Direction of Technical Change. *Quarterly Journal of Economics*, 120(2):517–549, 2005.

C. I. Jones. Recipes and economic growth: A combinatorial march down an exponential tail. *Working Paper*, 2022.

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

B. Lester, L. Visschers, and R. Wolthoff. Meeting technologies and optimal trading mechanisms in competitive search markets. *Journal of Economic Theory*, 155:1–15, 2015.

R. E. Lucas. Ideas and Growth. *Economica*, 76(301):1–19, 2009.

S. Mangin. When is competition price-increasing? the impact of expected competition on prices. *Working paper*, 2022.

E. Oberfield. A theory of input–output architecture. *Econometrica*, 86(2):559–589, 2018.

S. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987.

9 Appendix: Proofs

9.1 Proof of Lemma 1

By using (2), it is easy to verify that *(i)* holds by letting $y = 1$, and *(iii)* holds by assumption. Next, we have

$$\mathbb{P}_0'(z) = -\frac{1}{\theta}\sum_{n=0}^{\infty} n\mathbb{P}_n(\theta)\left(1 - \frac{z}{\theta}\right)^{n-1}. \tag{82}$$

Setting $z = 0$, this implies $\mathbb{P}_0'(0) = -1$ because $\sum_{n=0}^{\infty} n\mathbb{P}_n(\theta) = \theta$, so *(ii)* holds. ■

9.2 Proof of Lemma 3

If $\mathbb{P}_n$ satisfies **A2**, then $\mathbb{P}_0$ is continuous and infinitely differentiable. Moreover, as shown in Cai et al. (2022), we have $(-1)^k\mathbb{P}_0^{(k)}(z) \geq 0$ for all $k \in \mathbb{N}$ and $z \in \mathbb{R}^+$. Therefore, $\mathbb{P}_0$ is a completely monotone function. By the Bernstein-Widder theorem, there exists a Laplace transform representation, $\mathbb{P}_0(\theta) = \int e^{-\theta t}dF(t)$ for some finite measure F on the positive reals. Using the fact that $\mathbb{P}_0(0) = 1$ from Lemma 1, it follows that F must be a probability measure. Also, the mean of F must be one. To see this, we have

$$\mathbb{P}_0^{(k)}(\theta) = (-1)^k \int t^k e^{-\theta t} dF(t) \tag{83}$$

where $\mathbb{P}_0^{(k)}(0)$ is the $n-$th derivative of $\mathbb{P}_0$, evaluated at zero. Setting $k = 1$, the resulting expression evaluated at $\theta = 0$ gives us $\mathbb{P}_0'(0) = -\int tdF(t)$. Finally, Lemma 1 implies $\mathbb{P}_0'(0) = -1$, so $\int tdF(t) = 1$, i.e. $\mathbb{E}_F(x) = 1$. ■

9.3 Proof of Corollary 1

We know from Lester et al. (2015) that $\mathbb{P}_n$ is invariant if and only if

$$\mathbb{P}_n(\theta) = \frac{(-1)^n\theta^n\mathbb{P}_0^{(n)}(0)}{n!}. \tag{84}$$

Using (83) for the $k-$th derivative, and setting $k = n$ and $\theta = 0$, we obtain (25). ■

9.4 Proof of Lemma 4

Applying Corollary 1, the characteristic function of $N(\theta)/\theta$ is equal to

$$\begin{aligned}\mathbb{E}\left(e^{itN(\theta)/\theta}\right) &= \sum_{n=0}^{\infty} e^{itn/\theta}\mathbb{P}_n(\theta) \\ &= \sum_{n=0}^{\infty} e^{itn/\theta}\int_0^{\infty} e^{-\theta x}\frac{(\theta x)^n}{n!}dF(x) \\ &= \int_0^{\infty} e^{\theta x(e^{it/\theta}-1)}dF(x)\end{aligned}$$

Note that we have

$$\begin{aligned}\lim_{\theta\to\infty}\theta(e^{it/\theta}-1) &= \lim_{\theta\to\infty}\theta\left(\sum_{k=0}^{\infty}\frac{(it/\theta)^k}{k!}-1\right) \\ &= \lim_{\theta\to\infty}\theta\left(1+\frac{it}{\theta}+\sum_{k=2}^{\infty}\frac{(it/\theta)^k}{k!}-1\right) \\ &= \lim_{\theta\to\infty}\left(it+\sum_{k=2}^{\infty}\frac{(it)^k/\theta^{k-1}}{k!}\right) \\ &= it.\end{aligned}$$

Thus,

$$\begin{aligned}\lim_{\theta\to\infty}\mathbb{E}\left(e^{itN(\theta)/\theta}\right) &= \lim_{\theta\to\infty}\int_0^{\infty} e^{\theta x(e^{it/\theta}-1)}dF(x) \\ &= \int_0^{\infty} e^{xit}dF(x) \\ &= \mathbb{E}_F(e^{Xit}).\end{aligned}$$

This shows that the limit of the characteristic function of $N(\theta)/\theta$ is equal to the characteristic function of a random variable with distribution F. Thus, by Lévy's continuity theorem, the distribution of $N(\theta)/\theta$ converges to F. ■

9.5 Proof of Theorem 1

To prove Theorem 1, we use the following theorem from Barndorff-Nielsen (1964).

Theorem 5. *Let a_n and b_n be sequences of normalizing constants such that $\mathbb{P}(a_n M_n + b_n \leq x)$ converges to $H_\gamma(x)$. Let $N(n)$ be a sequence of discrete random variables such that $N(n)/n$ converges in probability to some random variable τ, and suppose $\mathbb{P}(\tau \leq 0) = 0$. Then,*

$$\lim_{n\to\infty} \mathbb{P}(a_n M_{N(n)} + b_n \leq x) = \int_0^\infty H_\gamma(x)^t d\mathbb{P}(\tau \leq t). \tag{85}$$

If $\mathbb{P}_n$ satisfies **A2**, then by Lemma 4, we know that $N(\theta)/\theta$ converges in distribution to a random variable X with distribution F equal to the mixing distribution of the mixed Poisson representation of $\mathbb{P}_n$ given by Lemma 3.

By Skorokhod's representation theorem, we can assume that $N(\theta)/\theta$ converges *in probability* to a random variable τ with distribution F. Also, we have

$$\mathbb{P}(X \leq 0) = \lim_{\theta\to\infty} \mathbb{P}(N(\theta)/\theta \leq 0) = \lim_{\theta\to\infty} \mathbb{P}_0(\theta), \tag{86}$$

and $\lim_{\theta\to\infty} \mathbb{P}_0(\theta) = 0$ by assumption. Thus, the conditions of Theorem 5 apply, so

$$\lim_{\theta\to\infty} \mathbb{P}(a_\theta M_{N(\theta)} + b_\theta \leq x) = \int_0^\infty H_\gamma(x)^t dF(t), \tag{87}$$

which can be written as follows:

$$\int_0^\infty H_\gamma(x)^t dF(t) = \int_0^\infty e^{\ln H_\gamma(x)^t} dF(t). \tag{88}$$

Recall that Lemma 3 says that if $\mathbb{P}_n$ satisfies **A2**, the function $\mathbb{P}_0$ can be written as

$$\mathbb{P}_0(\theta) = \int e^{-\theta t} dF(t) \tag{89}$$

where F is the mixing distribution of $\mathbb{P}_n$. Therefore, we have

$$\begin{aligned} H_{\gamma,\mathbb{P}}(x) &= \int_0^\infty H_\gamma(x)^t dF(t) && (90)\\ &= \int_0^\infty e^{-(-\ln H_\gamma(x))t} dF(t) && (91)\\ &= \mathbb{P}_0(-\ln(H_\gamma(x))). \end{aligned}$$

Finally, we are assuming that $H_\gamma(x) = e^{-v_\gamma(x)}$. Therefore, we obtain

$$H_{\gamma,\mathbb{P}}(x) = \mathbb{P}_0(v_\gamma(x)), \tag{92}$$

which completes the proof. ■

9.6 Proof of Theorem 2

It follows from Assumption **A2** that the associated pdf of $H_{\mathbb{P}}(x;\theta)$ is

$$h_{\mathbb{P}}(x;\theta) = \frac{-g(x)\theta\mathbb{P}_0'(\theta(1-G(x)))}{1-\mathbb{P}_0(\theta)} \tag{93}$$

for $(\underline{x}, \bar{x})$. Defining $k : \mathbb{R}^+ \to [0,1]$ by $k(z) = -\mathbb{P}_0'(z)$, we have

$$h_{\mathbb{P}}(x;\theta) = \frac{g(x)\theta k(\theta(1-G(x)))}{1-\mathbb{P}_0(\theta)}. \tag{94}$$

Therefore, given that $\lim_{\theta\to\infty}\mathbb{P}_0(\theta) = 0$, we have

$$\mathbb{E}_{H_{\mathbb{P}}}(\zeta(x)) \sim_{\theta\to\infty} \int_{\underline{x}}^{\infty} \zeta(x)g(x)\theta k(\theta(1-G(x)))dx. \tag{95}$$

Consider the integral on the right-hand side of (95). For any given θ, this integral is finite because **A3** says that $\int_{\underline{x}}^{\bar{x}} |\zeta(x)g(x)|\,dx < \infty$ and we have $k : \mathbb{R}^+ \to [0,1]$. Changing variables in the integral in (95) using $x = G^{-1}\left(1-\frac{1}{t}\right)$, we obtain

$$\int_{\underline{x}}^{\bar{x}} \zeta(x)h_{\mathbb{P}}(x;\theta)dx = \int_1^{\infty} k\left(\frac{\theta}{t}\right)\frac{\theta}{t}\zeta\left(G^{-1}\left(1-\frac{1}{t}\right)\right)\frac{dt}{t}. \tag{96}$$

Define $k_0 : \mathbb{R}^+ \to \mathbb{R}^+$ by $k_0(z) := zk(z)$ 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:

$$\int_{\underline{x}}^{\bar{x}} \zeta(x)h_{\mathbb{P}}(x;\theta)dx = \int_0^{\infty} k_0\left(\frac{\theta}{t}\right)\widetilde{\zeta}(t)\frac{dt}{t}. \tag{97}$$

Rewriting the above, we have

$$\int_{\underline{x}}^{\bar{x}} \zeta(x)h_{\mathbb{P}}(x;\theta)dx = (k_0 \overset{M}{*} \widetilde{\zeta})(\theta), \tag{98}$$

where $(\phi \overset{M}{*} f)(\theta)$ denotes the Mellin convolution of ϕ and f, evaluated at θ, defined by

$$(\phi \overset{M}{*} f)(\theta) := \int_0^\infty \phi\left(\frac{\theta}{t}\right) f(t)\frac{dt}{t}. \tag{99}$$

We can now apply Theorem 4.1.6 from (Bingham et al., 1987, §4.0.1) . This theorem says that if *(i)* $f : [0,\infty) \to \mathbb{R}$ is measurable, *(ii)* $f(t)/t^\sigma$ is bounded on $(0, t]$ for any $t > 0$, *(iii)* $f(t) \in RV_{-\upsilon}^{\infty}$, and *(iv)* there exists $\sigma, \tau \in \mathbb{R}$ such that $\upsilon \in (\sigma, \tau)$ and, for all $s \in [\sigma, \tau]$, the Mellin transform $\breve{\phi}(-s) := \int_0^\infty t^{-s-1}\phi(t)dt$ is finite, then

$$(\ \phi \overset{M}{*} f)(\theta) \sim_{\theta\to\infty} \breve{\phi}(-\upsilon)f(\theta). \tag{100}$$

We verify that the conditions for applying this theorem hold when $f = \widetilde{\zeta}$, $\upsilon = -\rho$, and $\phi = k_0$. *(i)* By assumption **A3**, ζ is measurable, so $\widetilde{\zeta}$ is measurable. *(ii)* Since $\widetilde{\zeta}(t) = 0$ for all $t < 1$, this 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 $\widetilde{\zeta}$, this is in turn equivalent to boundedness of ζ on every interval of the form $[\underline{x}, x]$, which is true by assumption **A3**. *(iii)* By assumption, $\zeta(G^{-1}(1-t)) \in RV_\rho^0$, which implies $\widetilde{\zeta}(t) \in RV_{-\rho}^{\infty}$ since $\widetilde{\zeta}(t) = \zeta(G^{-1}(1 - 1/t))$ for $t \in [1,\infty)$. *(iv)* By Lemma 5, the Mellin transform $\breve{k}_0(a)$ equals $\Gamma(a)E_F(X^a)$, which is finite if $a > -1$ and $E_F(X^a)$ is finite.

We can now apply Theorem 4.1.6 from (Bingham et al., 1987, Theorem 4.1.6) to obtain

$$(k_0 \overset{M}{*} \widetilde{\zeta})(\theta) \sim_{\theta\to\infty} \breve{k}_0(\rho)\widetilde{\zeta}(\theta). \tag{101}$$

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

$$\int_{\underline{x}}^{\bar{x}} \zeta(x)h_{\mathbb{P}}(x;\theta)dx \sim_{\theta\to\infty} \zeta\left(G^{-1}\left(1 - \frac{1}{\theta}\right)\right)\int_0^\infty t^\rho k(t)dt. \tag{102}$$

Using $k(t) = -\mathbb{P}_0'(t)$, we have $\int_0^\infty t^\rho k(t)dt = -\int_0^\infty t^{a-1}\mathbb{P}_0'(t)dt$, which yields (40). ■

9.7 Proof of Lemma 5

By definition, we have

$$\Psi_{\mathbb{P}}(a) = -\int_0^\infty t^{a-1}\mathbb{P}_0'(t)dt. \tag{103}$$

From Lemma 3, we obtain

$$-\int_0^\infty t^{a-1}\mathbb{P}_0'(t)dt = \int_0^\infty t^{a-1}\int_0^\infty ue^{-tu}dF(u)dt. \tag{104}$$

We perform the change of variables $v = tu$ to get

$$\begin{aligned} \int_0^\infty t^{a-1}\int_0^\infty ue^{-tu}dF(u)dt &= \int_0^\infty (v/u)^{a-1}\int_0^\infty e^{-v}dF(u)dv \\ &= \int_0^\infty v^{a-1}e^{-v}dv\int_0^\infty u^{1-a}dF(u) \\ &= \Gamma(a)\mathbb{E}_F(X^{1-a}), \end{aligned} \tag{105}$$

which completes the proof. ■