

Federal Reserve Bank of Minneapolis  
Research Department

## Models of Growth and Firm Heterogeneity\*

Erzo G.J. Luttmer

Working Paper 678

April 2010

### ABSTRACT

---

Although employment at individual firms tends to be highly non-stationary, the employment size distribution of all firms in the United States appears to be stationary. It closely resembles a Pareto distribution. There is a lot of entry and exit, mostly of small firms. This paper surveys general equilibrium models that can be used to interpret these facts and explores the role of innovation by new and incumbent firms in determining aggregate growth. The existence of a balanced growth path with a stationary employment size distribution depends crucially on assumptions made about the cost of entry. Some type of labor must be an essential input in setting up new firms.

---

\*Luttmer, University of Minnesota and Federal Reserve Bank of Minneapolis. Prepared for the Annual Reviews of Economics 2010. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

## 1. INTRODUCTION

It has long been known that the employment size distribution of firms is highly skewed. In modern US data, the number of firms with more than  $n$  employees behaves roughly like  $1/n^\zeta$ , for some tail index  $\zeta$  slightly greater than 1. This implies that a large share of aggregate employment is accounted for by a relatively small number of large firms. There are about 6 million employer firms in the US. Around half of the labor force of these firms is employed by the roughly 18,000 firms with more than 500 employees, and a good quarter is accounted for by the 1,000 or so firms with more than 10,000 employees. This pattern of extreme skewness appears to be quite stable over time, although the existence of large private corporations is of course a relatively modern phenomenon.

Firms play an important role in theories of aggregate growth and fluctuations. Yet, few studies in which firm boundaries matter are consistent with the highly skewed size distribution observed in the data. The situation is different in the modern trade literature. Helpman, Melitz and Yeaple [2004], Melitz [2003], Bernard, Eaton, Jensen and Kortum [2003], Eaton, Kortum and Kramartz [2008] are leading examples of quantitative theories of trade that explicitly take into account the firm size distribution. In most of the work on trade, the Fréchet or Pareto-like size distribution is a direct reflection of an underlying productivity distribution. But the origin of this productivity distribution is typically left unexplained, and one could argue that this amounts to setting aside the most important puzzle in the data. Even the extreme-value interpretation of Kortum [1997] and Eaton and Kortum [1999] depends on an underlying distribution of productivities that is itself skewed. In large samples, the maximum of a random sample from a distribution with compact support does not converge to a thick-tailed Fréchet distribution but piles up near the upper bound of that support.<sup>1</sup>

This paper surveys general equilibrium models that can be used to interpret the  $1/n^\zeta$  phenomenon. These models build on a rich history of research on firm growth and size distributions. No attempt is made here to cover this history (see Sutton [1997], Neal and Rosen [2000], and Gabaix [1999, 2009].) Although several examples are taken from the trade literature, trade is not the focus here. This survey also cannot do justice to the vast and related literature on innovation and growth. Excellent survey and textbook presentations can be found in Aghion and Durlauf [2005] and Acemoglu [2008]. The object of interest here is the firm—not establishments, such as manufacturing plants,

---

<sup>1</sup>See Alvarez, Buera and Lucas [2008] for more on the relation between the distribution from which productivities are drawn and the distribution of frontier productivities.

administrative offices, or retail stores.<sup>2</sup> In data constructed by the U.S. Census, there is a fairly clear empirical definition of an employer firm. Models differ in their interpretation of this empirical construct.

The discussion is organized around two polar interpretations of the firm size distribution. One is that the  $1/n^\zeta$  tail is the result of replication of organization capital. A firm is defined by its firm-specific organization capital, and it takes pieces of this capital to create more pieces. This can easily give rise to Gibrat's law—the proposition that firm growth rates are independent of size. If there is also persistent entry of small new firms, then the Pareto-like tail follows. In the simplest version of this interpretation, all organization capital is the same and there are no productivity differences. All measured productivity differences must be attributed to unobserved differences in the quantity of organization capital across firms.<sup>3</sup> More plausibly, there are differences in the quality of organization capital across firms. Firms with high-quality organization capital have strong incentives to create more of it. These firms will grow fast and large firms will be those that have had frontier-quality organization capital for a sufficiently long time.

The second polar interpretation is that firm size directly reflects productivity differences, moderated only by decreasing returns to scale or downward sloping demand curves. The scale of the firm can be adjusted instantaneously and firm growth is the result of productivity growth. If productivity growth is independent of productivity levels, Gibrat's law can again arise and entry of sufficiently productive new firms will result in a stationary distribution with a Pareto-like tail. Randomness in firm-level productivity growth can be interpreted as learning by doing or experimentation. Firms with sufficiently many lucky draws survive and others are forced to exit. If entrants can do better than these exiting firms, the economy-wide productivity distribution will trend upwards over time.

In both types of economies, not enough entry would result in non-stationary size distributions, typically with a thin right tail. When the productivity of incumbent firms improves over time, enough entry to induce stationarity can only occur if potential entrants can take advantage of the improvements made by incumbents.<sup>4</sup> This can happen

---

<sup>2</sup>See Rossi-Hansberg and Wright [2007] for empirical evidence and a general equilibrium model of establishment size dynamics. In the US, the right tail of the size distribution of establishments is noticeably thinner than that of firms.

<sup>3</sup>McGrattan and Prescott [2009] emphasize the importance of unmeasured investment.

<sup>4</sup>An important alternative possibility, not considered here, is that relative prices change to allow new industries to arise. Entrants into these new industries need not learn anything from incumbents in old industries.

through trade of something that embodies those improvements, or as a result of imitation externalities. The welfare implications are different. Both mechanisms for the transfer of knowledge are likely to play a role empirically, and it is a continuing challenge to quantify their relative importance. The fact that the stationary distribution has a tail index  $\zeta$  only slightly above 1 arises because, while there is sufficient entry for stationarity, incumbent firms account for much of aggregate employment growth.

The rest of this survey is divided into three parts. The motivating evidence is briefly discussed in Section 2. The organization capital and productivity interpretations are presented in Sections 3 and 4, with hybrid models appearing in both sections. Throughout, the focus is on examples that are sufficiently tractable to allow one to think through aggregate implications.<sup>5</sup>

## 2. SOME MORE FACTS

Figure 1 shows that the aggregate civilian labor force and various firm counts exhibit a common trend over the past 80 or so years. Completely consistent firm counts over this period are not available. Lucas [1978] used the series V13 “Firms in Operation” published in the Statistical Abstract of the United States. The County Business Patterns (CBP) publication of the U.S. Census provides a long series of what were referred to as reporting units before 1974 and subsequently became establishments. The Small Business Administration (SBA) reports both a count of firms and of establishments, including establishments that have zero employment during the March reporting period. Establishments that do have employment in that period correspond to the establishments reported in the Business Dynamics Statistics (BDS) of the Census, available since 1977. The old V13 firm count was discontinued after 1963 after errors were discovered. Rather mysteriously, the V13 number of firms exceeds the County Business Patterns number of reporting units by a wide margin. As one would expect, the SBA establishment count exceeds its firm count.

The firm size distributions reported by the SBA for 1992, 2000 and 2006 are shown in Figure 2. The first and second panel show the left and right cumulative distribution functions, respectively. The Pareto right tail and the fact that the tail index is close to 1 are evident. A simple regression based on firms with more than 100 employees gives a

---

<sup>5</sup>See Weintraub, Benkard and van Roy [2008] for an attempt to narrow the gap between the Hopenhayn [1992] style models considered here and models that allow for dynamic strategic interaction such as Ericson and Pakes [1995].

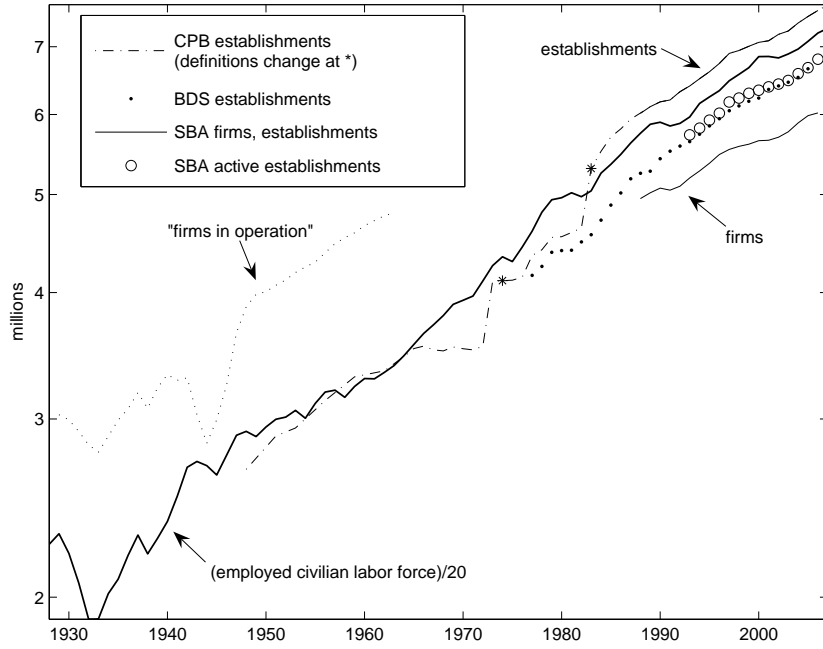


FIGURE 1 Firm Counts and Aggregate Employment

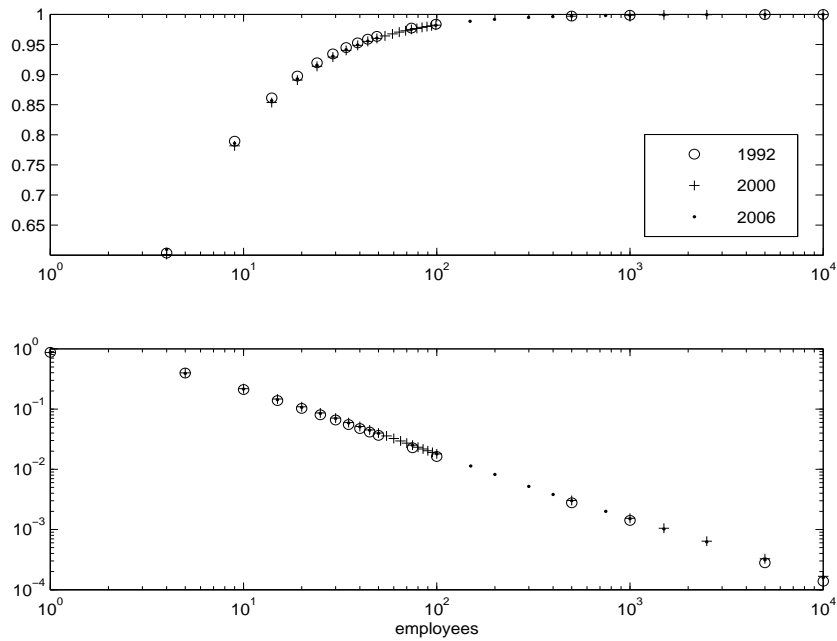


FIGURE 2 The Firm Size Distribution

tail index of  $\zeta \approx 1.06$ . Although the SBA data do not go back in time very far, Figure 2 clearly suggests a size distribution that is stationary. This stability can be illustrated further using BDS data that go back to 1977. The BDS does not show firm counts but it does report aggregate employment accounted for by various firm size categories. The

share of aggregate employment in the right tail of the firm size distribution as reported by the BDS is shown in Figure 3.

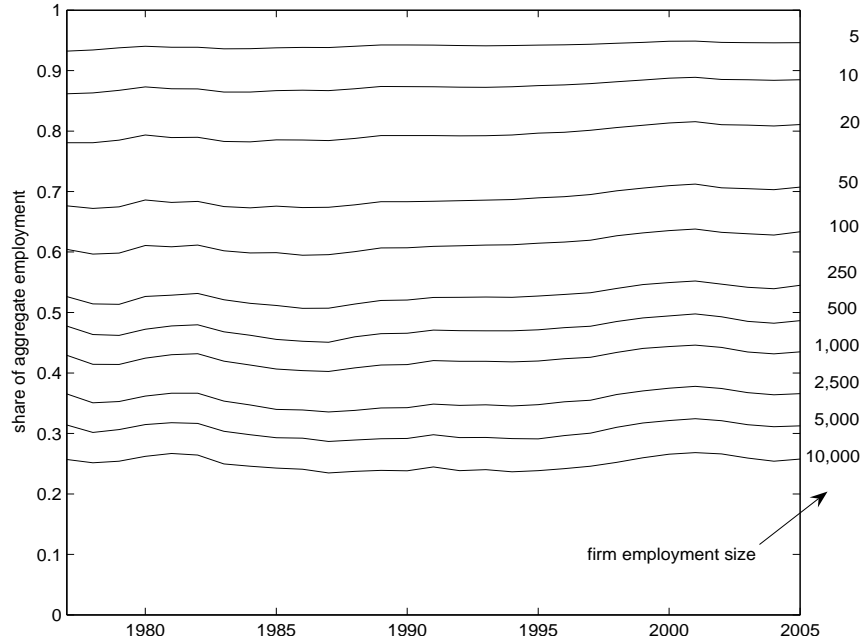


FIGURE 3 The Right Cumulative Distribution of Employment Across Firms

Although the repeated cross-sections are stable, the picture for individual firms could not be more different. Since Gibrat [1931], the empirical benchmark is that firms grow according to Gibrat's law: growth rates are independent of size. Thus the size of an individual firm is non-stationary. Sutton [1997] surveys the literature. Hall [1987] suggests that Gibrat's law cannot be rejected for large firms and only weakly for small firms. Evans [1987] finds that firm growth rates decrease with size, even after controlling for selection issues related to exit. He also finds an important role for age, as do Dunne, Roberts and Samuelson [1989]. Small and young firms are very volatile and highly likely to exit. SBA data for 1988-2006 show an average exit rate of about 10.4% per annum for firms with fewer than 20 employees, versus 2.5% for firms with 500 or more employees. Compared with small firms, few large and established firms disappear in a given year.

A revealing piece of evidence on what large firms are like is shown in Figure 4. It shows the average number of establishments of all firms in the firm employment size categories reported by the SBA. Not shown is the average number of establishments per firm, equal to 676, in the unbounded size category of 10,000 or more employees. For firms with more than about 100 employees, the relationship is essentially log-linear, with

an elasticity only slightly below 1. The average large firm has many establishments, and not, typically, a single large establishment.

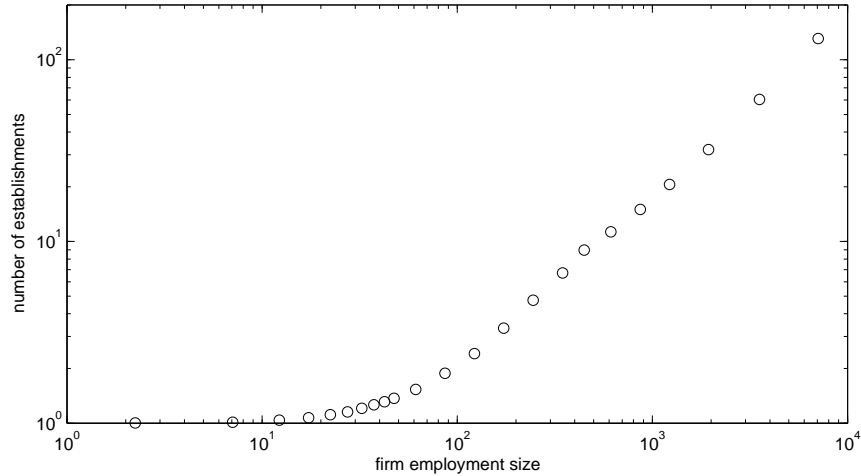


FIGURE 4 The Average Number of Establishments per Firm

### 3. ORGANIZATION CAPITAL

A natural way to obtain Gibrat's law and Pareto-like size distributions is to take firm size to reflect accumulated organization capital, as in Prescott and Visscher [1980]. In the examples given here, entrepreneurs create start-up capital, and this capital can be used to produce consumption and more of the same capital. A firm is identified with all the capital produced, directly or indirectly, from the same initial start-up stock of capital. Transferring capital from one firm to another is taken to be sufficiently costly for it not to occur.

#### 3.1 A Deterministic Example

Consider an economy with a population of infinitely lived infinitesimal agents  $H_t = He^{\eta t}$  that grows at a positive rate. Agents have a subjective discount rate  $\rho > \eta$  and logarithmic utility over per-capita consumption flows. Everyone can exert one unit of effort per unit of time. This effort can be used to supply labor or to create start-up capital, as an entrepreneur. Individuals are endowed with a skill vector that determines the amount of labor they can supply, and the rate at which they can create start-up capital. The distribution of these skills is time invariant, and agents choose to be workers or entrepreneurs based on comparative advantage, as in Roy [1951]. Specifically, suppose the price of a unit of start-up capital is  $q_t$  and the wage is  $w_t$ , both in units

of consumption. An agent who can supply  $x$  units of labor or create a unit of start-up capital at a Poisson rate  $y$  will choose to be a worker if  $w_t x > q_t y$  and an entrepreneur if the reverse inequality holds. Using the skill distribution to sum over all agents will give rise to a per-capita labor supply  $L(q_t/w_t)$  and a per-capita supply of start-up capital  $E(q_t/w_t)$ . Clearly, these functions are decreasing and increasing in  $q_t/w_t$ , respectively.

A newly created firm starts out with one unit of start-up capital. Once the firm is created, there is joint production: its capital can simultaneously be used to produce consumption goods and more capital. Consumption can be produced according to a production function  $F(n_t, n_t l_t)$ , where  $n_t \in [1, \infty)$  is the capital stock of the firm, and  $n_t \times l_t$  is production labor. The firm's capital stock grows according to

$$Dn_t = G(n_t, n_t m_t) \quad (1)$$

where  $n_t m_t$  is capital-producing labor. Both  $F$  and  $G$  are increasing, concave, and exhibit constant returns to scale. The assumption of constant returns ensures that the value of the firm can be written as the price of a unit of capital times the capital stock of the firm. Given an interest rate  $r_t$ , the price of one unit of capital satisfies the Bellman equation

$$r_t q_t = \max_{l, m \geq 0} \{F(1, l) + q_t G(1, m) - w_t(l + m) + Dq_t\} \quad (2)$$

and a transversality condition.

This structure is clearly reminiscent of Lucas [1967] and Hayashi [1982]. But observe that final output in this economy is used only for consumption. Capital is produced either using old capital and labor, or from scratch by entrepreneurs. The economy has a balanced growth path in which per-capita consumption is constant and capital is accumulated to keep up with population growth. Along the balanced growth path, the interest rate is  $r_t = \rho$  and  $[q_t, w_t] = [q, w]$ . Given interest rates and wages, the price of a unit of capital is determined by

$$w = D_2 F(1, l) = q D_2 G(1, m), \quad q = \frac{F(1, l) - w(l + m)}{\rho - G(1, m)}. \quad (3)$$

Firm employment equals  $l + m$  times the capital stock of the firm. Let  $N e^{nt}$  denote the aggregate capital stock along the balanced growth path. The labor market clearing condition is then

$$(l + m)N = L(q/w)H. \quad (4)$$

Existing capital produces new capital at the rate  $G(1, m)$ . Entrepreneurs account for any additional growth in the capital stock. Since both aggregate capital and the population



of entrepreneurs grow at the rate  $\eta$ , it must be that

$$\eta N = G(1, m)N + E(q/w)H. \quad (5)$$

The balanced growth path is determined by solving (3)-(5) for  $N$ ,  $l$ ,  $m$ ,  $q$  and  $w$ . Assuming capital is essential for producing consumption,  $N$  will be positive. Together with  $E(q/w) \geq 0$  this implies  $\eta \geq G(1, m)$ , and then the assumption  $\rho > \eta$  ensures  $\rho > G(1, m)$ .

Suppose first that the talent distribution is such that  $E(q/w) = 0$  for low enough  $q/w$ . That is, the supply of start-up capital dries up when the price of capital is low enough. Then the balanced growth path may exhibit no investment in start-up capital by entrepreneurs. Incumbent firms grow at the rate  $\eta$  and  $N$  is determined by initial conditions. Although there is balanced growth, the firm size distribution is non-stationary: the distribution of log capital shifts to the right as incumbent firms produce new capital at a rate  $\eta$ .

Suppose instead that  $E(q/w)$  is strictly positive for all  $q/w$  positive. There are always some entrepreneurs for whom it is profitable to create new start-up capital. Then (5) implies that the balanced growth path must be such that  $G(1, m) = \mu < \eta$ . A natural assumption is that  $G(1, 0)$  is negative, and so  $\mu$  need not be positive in equilibrium. Firms could enter and then shrink forever. Throughout the following, assume this is not the case. Because  $\mu < \eta$ , there is continuous entry of new firms. The number of firms, as well as the number of entering firms, grows at the rate  $\eta$ . At time  $t$  the number of firms in the cohort of firms of age  $a$  is proportional to  $e^{\eta(t-a)}$ . The age distribution of firms is therefore stationary, with a density  $\eta e^{-\eta a}$ . As new firms all enter with one unit of capital and accumulate capital at the common rate  $\mu$ , a firm of age  $a$  has  $n_a = e^{\mu a}$  units of capital. The size of a firm is a deterministic function of age. If  $\mu$  is positive, then changing variables from age to capital gives a distribution of firm size in terms of capital that has a density<sup>6</sup>

$$p(n) = \frac{\eta}{\mu} n^{-(1+\eta/\mu)},$$

for all  $n \in [1, \infty)$ . The firm size distribution is Pareto. Since employment and output scale with capital, this is true for any of these measures of firm size.

The right-tail probabilities of  $p(n)$  are  $n^{-\zeta}$ , where  $\zeta = \eta/\mu$ . This implies the log-linear relationship shown in Figure 2, and its slope reveals the tail index  $\zeta$  of the distribution. Because (5) implies  $\eta > \mu$  when entry is positive, the tail index  $\zeta$  is guaranteed to be

---

<sup>6</sup>Benhabib and Bisin [2006] trace this argument back to Cantelli and Fermi.

above 1 if at any price there will be some entrepreneurs who choose to create start-up capital. This ensures the size distribution has a finite mean, given by  $\zeta/(\zeta - 1)$ . Zipf's law arises when  $\zeta$  approaches 1 from above. But this limiting distribution is not an equilibrium distribution in this economy. As seen above, without entry the size distribution is non-stationary.

The fraction of all capital held by firms with more than  $n$  units of capital is  $n^{-(\zeta-1)}$ , which approaches 1 for any  $n$  as the equilibrium distribution approaches Zipf's law. If the smallest firm has one employee, then the fact that firms with more than 500 employees account for 50% of employment implies  $\zeta = 1 + \ln(2)/\ln(500) = 1.11$ . An analogous calculation for the 27% of employment accounted for by firms with more than 10,000 employees yields the very similar  $\zeta = 1.14$ , a reflection of the accuracy of the fit of the Pareto distribution in this range of the data.

The formula  $\zeta = \eta/\mu$  implies that  $\mu$  must be close to the population growth rate. In the United States this is approximately 1% per annum, and so firms in this economy should grow at a rate of about .9% per annum. If new firms enter with one employee, then one obtains the rather problematic implication that it takes  $\ln(10,000)/.009 = 1,023$  years to become one of the roughly 1,000 firms with more than 10,000 employees observed in US data.

A further anomalous implication of this simple model is that there is no exit. In U.S. data the exit rate is approximately 10% per annum. An easy fix is to assume that firms die randomly, with all their capital, at a rate  $\delta$ . The equilibrium conditions (3)-(5) require only minor modification:  $\rho$  must be replaced by  $\rho + \delta$  in (3) and  $\eta$  by  $\eta + \delta$  in (5). The number of firms of age  $a$  at time  $t$  is proportional to  $e^{\eta(t-a) - \delta a}$ , and so the age distribution has a density  $(\eta + \delta)e^{-(\eta + \delta)a}$ . The resulting size distribution is Pareto with a tail index  $\zeta = (\eta + \delta)/\mu$ . Population growth is no longer needed to obtain a Pareto size distribution. What matters is that firms have a chance to grow. The resulting distribution will fit the empirical size distribution if  $\mu$  is a little below  $\eta + \delta$ . Now we can have  $\mu \approx .1$ , and then firms that start with one employee only take  $\ln(10,000)/.1 = 92$  years to reach 10,000 employees. This is closer to the median age of 75 years for firms of 10,000 or more employees reported in Luttmer [2008]. But, as reported in Section 2, the assumption that the likelihood of exit is independent of firm size is starkly at odds with US data.

### 3.2 A Brownian Example

Firm age and logarithmic employment size are positively correlated in US data. But the correlation is far from perfect. A simple way to account for this is to introduce a random component to firm growth. Specifically, assume the capital stock of a particular firm evolves according to

$$dn_t = n_t [G(1, m_t)dt + \sigma dW_t], \quad (6)$$

where  $W_t$  is a firm-specific standard Brownian motion and  $m_t$  is the amount of labor used to produce new capital, per unit of existing capital. The Brownian component of (6) amounts to a common multiplicative shock to all the units of capital inside the firm. This is still a technology that exhibits constant returns to scale. But firm boundaries matter, in the sense that re-allocating a unit of capital from one firm to another exposes that unit of capital to different shocks.

Along a balanced growth path,  $r_t = \rho$ ,  $w_t = w$ , and the value of a firm with  $n$  units of capital,  $Q(n)$ , must satisfy the Bellman equation

$$\rho Q(n) = \max_{l, m} \left\{ F(n, nl) - wn(l + m) + G(n, nm)DQ(n) + \frac{1}{2}\sigma^2 n^2 D^2 Q(n) \right\}.$$

The constant-returns-to-scale assumptions imposed on  $F$  and  $G$  immediately suggest a solution of the form  $Q(n) = qn$ . With this conjecture, the Bellman equation reduces to the balanced growth version of the deterministic Bellman equation (2). The result is a constant  $m_t = m$ , and this implies that firm growth satisfies Gibrat's law, in a strict sense: for any  $\Delta > 0$ , the distribution of  $n_{t+\Delta}/n_t$  is independent of firm size  $n_t$ , and of anything else. Conditions (3)-(5) continue to define the balanced growth path.

To compute the size distribution, apply Ito's lemma to (6) to conclude that the drift of  $\ln(n_t)$  is equal to  $\mu = G(1, m) - \sigma^2/2$ . The log-size distribution of a cohort of age  $a$  is normal with mean  $\mu a$  and variance  $\sigma^2 a$ . If the balanced growth path is such that  $\eta = G(1, m)$ , then there is no entry. If all firms are initially the same, then the size distribution along the balanced growth path is log normal with mean  $(N/H) + \mu t$  and variance  $\sigma^2 t$ . There is no stationary size distribution in that case. In contrast to the deterministic example, scaling firm size with an exponential trend does not make the size distribution stationary.

If  $G(1, m) < \eta$  then there is positive entry and the age distribution will have a density  $\eta e^{-\eta a}$  along the balanced growth path. Integrating the normal density for log size given age against this exponential age density and converting the resulting density for log size

back to size gives what is known as the double Pareto distribution<sup>7</sup>

$$p(n) = \frac{\zeta\zeta_*}{\zeta + \zeta_*} \min \{n^{\zeta_*-1}, n^{-(1+\zeta)}\}, \quad (7)$$

where

$$\zeta = -\frac{\mu}{\sigma^2} + \sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + \frac{\eta}{\sigma^2/2}}, \quad \zeta_* = \frac{\mu}{\sigma^2} + \sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + \frac{\eta}{\sigma^2/2}}. \quad (8)$$

Another way to show this result is to use the fact that the density  $f$  of  $s = \ln(n)$  satisfies the Kolmogorov forward equation  $\eta f(s) = -\mu Df(s) + \frac{1}{2}\sigma^2 D^2 f(s)$  for all  $s$ , except at the entry point  $s = 0$ . The density is continuous at that point, but there is a kink  $\eta = \frac{1}{2}\sigma^2[D_-f(0) - D_+f(0)]$  that reflects the entry that takes place at  $s = 0$ . The remaining boundary conditions are that  $f$  is positive and integrates to 1. Note that  $G(1, m) < \eta$  implies  $\mu + \sigma^2/2 < \eta$ . This inequality is equivalent to  $\zeta > 1$ , which is precisely the condition needed to ensure that the mean firm size,  $\zeta\zeta_*/(\zeta - 1)(\zeta_* + 1)$ , is finite. The tail index  $\zeta$  approaches 1 from above as  $\mu + \sigma^2/2$  increases towards  $\eta$ .

All this is predicated on  $\eta > 0$ . But, if there is also random exit at a rate  $\delta$ , then  $\eta$  must be replaced by  $\eta + \delta$  in the above calculations, and one only needs to assume  $\eta + \delta > 0$ .

### 3.3 Stochastic Replication

Taking organization capital to be continuous makes for easy calculations but does not aid interpretation. Organizations can be viewed as collections of matches, contracts, or trading relationships, that are more naturally taken to be discrete. One is an employee, supplier, or customer of a firm, or one is not. Large firms typically operate many plants, offices or stores that are geographically dispersed. In the following, the discrete units that make up organization capital are referred to generically as “blueprints.” Firm growth is about the replication of these blueprints.

#### 3.3.1 Independent Replication

Entrepreneurs create a start-up “blueprint” that defines a new firm. This blueprint can be replicated within the firm, and every resulting blueprint can itself be replicated inside the firm. Let  $n_t \in \mathbb{N}$  denote the number of blueprints of the firm. Consumption produced

---

<sup>7</sup>See Reed [2001], Mitzenmacher [2004], and references therein. This uses the fact that for any  $A \geq 0$ ,  $\int_0^T \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t}(t - A)^2\right) dt = \Phi\left(\frac{-A+T}{\sqrt{T}}\right) - e^{2A}\Phi\left(\frac{-A-T}{\sqrt{T}}\right)$ , where  $\Phi$  is the standard normal cumulative distribution function. A recent application can be found in Arkolakis [2009].

by a firm with  $n_t$  blueprints is again  $F(n_t, n_t l_t)$ , where  $l_t$  is production labor used per blueprint. Combining  $m_t$  units of labor with a blueprint generates a new blueprint at a Poisson rate  $G(1, m_t)$ . In this example,  $G$  will be assumed to be strictly positive. The Bellman equation is again (2) and the economy will have a balanced growth path defined by (3)-(5). Write  $\mu = G(1, m)$  for the rate at which blueprints are replicated.

Let  $\{p_{n,a}\}_{n \in \mathbb{N}}$  denote the probability distribution of the number of blueprints of a firm at age  $a$ . These probabilities satisfy the differential equations

$$Dp_{1,a} = -\mu p_{1,a}, \quad Dp_{n+1,a} = \mu n p_{n,a} - \mu(n+1)p_{n+1,a}, \quad n \in \mathbb{N}, \quad (9)$$

and the initial condition is  $p_{1,0} = 1$ . Clearly,  $p_{1,a} = e^{-\mu a}$  and one can then proceed to construct  $p_{n,a}$  inductively. This results in a geometric distribution of size given age,

$$p_{n,a} = e^{-\mu a} (1 - e^{-\mu a})^{n-1}, \quad n \in \mathbb{N}. \quad (10)$$

Observe that the mean of this distribution is  $e^{\mu a}$ , as expected. There may again be a balanced growth path with no entry if entrepreneurs stop creating new firms at sufficiently low blueprint prices. As in the Brownian example, the resulting size distribution is non-stationary, and there is no way to de-trend firm size to make it stationary. A stationary distribution arises if the entry rate is positive, so that  $\eta > \mu$ . The age distribution again has the density  $\eta e^{-\eta a}$ . Combining this with (10) gives

$$p_n = \int_0^\infty \eta e^{-\eta a} p_{n,a} da = \frac{\eta}{\mu} \int_0^\infty e^{-(1+\eta/\mu)b} (1 - e^{-b})^{n-1} db.$$

There are several ways to calculate this integral. One is to change variables to  $x = e^{-b}$  and recognize the result as the beta function with parameters  $1 + \eta/\mu$  and  $n$ . Another is to expand  $(1 - e^{-b})^{n-1}$  using the binomial formula and do the integration term by term. Probably the most direct way is to use repeated integration by parts. The result is

$$p_n = \frac{\eta \Gamma(n) \Gamma(1 + \eta/\mu)}{\mu \Gamma(n + 1 + \eta/\mu)} \quad (11)$$

for all  $n \in \mathbb{N}$ . Here,  $\Gamma$  is the gamma function, which specializes to  $\Gamma(x) = (x-1)!$  for integer values of  $x$ . The result (11) is due to Yule [1925] and Simon [1955]. The process that gives rise to (11) is widely known as the Yule process (Feller [1968], Karlin and Taylor [1975], or Ross [1996].) Integrating the mean  $e^{\mu a}$  of firm size given age against the age density  $\eta e^{-\eta a}$  shows that (11) must have a mean  $\eta/(\eta - \mu)$ . Stirling's formula implies that  $n! \sim n^{n+\frac{1}{2}} e^{-n}$  and thus  $p_n \sim n^{-(1+\eta/\mu)}$ . To compute the right-tail probabilities, note that they are  $(1 - e^{-\mu a})^{n-1}$  for a cohort of age  $a$ . Averaging against

the age density  $\eta e^{-\eta a}$  gives an integral just like the one that defines  $p_n$ . The resulting right-tail probabilities are

$$\sum_{k=n}^{\infty} \frac{\eta \Gamma(k) \Gamma(1 + \eta/\mu)}{\mu \Gamma(k + 1 + \eta/\mu)} = \frac{\eta \Gamma(n) \Gamma(\eta/\mu)}{\mu \Gamma(n + \eta/\mu)} \sim n^{-\eta/\mu}.$$

A Zipf plot for this distribution is log-linear for large  $n$ , with a slope  $-\eta/\mu < -1$ . Zipf's Law arises as  $\mu \uparrow \eta$ , which gives  $p_n = 1/[n(n + 1)]$  and right-tail probabilities equal to  $1/n$ .

**Stochastic Depreciation** The Yule process implies that firms can only grow. A natural generalization is to allow individual blueprints to depreciate in one-hoss-shay fashion at some rate  $\lambda$ , where  $\lambda$  may depend on labor assigned to “maintain” blueprints. A firm exits when it loses its last blueprint and therefore has no blueprints that can be replicated anymore. The conditions for a balanced growth path are similar to (3)-(4) and will imply that  $\eta > \mu - \lambda$  if there is positive entry. Firm size follows a simple birth-death process that is a special case of the more general time-dependent birth-death process studied in Kendall [1948]. Conditional on survival, the cohort size distribution is again geometric as in (10), but the variable  $1 - e^{-\mu a}$  in (10) must be replaced by  $(e^{(\mu-\lambda)a} - 1)/(e^{(\mu-\lambda)a} - \lambda/\mu)$ . Although the population of firms grows exponentially, exit is highly size-dependent: only the firms with one remaining blueprint can exit. This seems to be closer to US data than random exit. A characterization of the stationary size and age distributions (no longer exponential) can be found in Luttmer [2008]. The right-tail probabilities of the size distribution behave like  $n^{-\zeta}$  with  $\zeta = \eta/(\mu - \lambda)$  when  $\mu > \lambda$ . The thick right tail observed in the data arises when firms are expected to grow at a positive rate.

### 3.3.2 Synchronized Replication

In the case of the Yule process and its birth-death generalization with constant  $\mu$  and  $\lambda$ , replication and decay are independent across blueprints. All randomness in firm growth arises at the blueprint level. At the opposite extreme, consider a replication technology with the feature that the random event of replication is perfectly synchronized across all blueprints within the same firm. Specifically, suppose a synchronized replication event occurs at a Poisson rate  $G(n, nm)$  if a firm with  $n$  blueprints uses  $nm$  units of labor to attempt replication. The arrival of such a synchronized replication event generates  $n$  new blueprints, and the cost of replication is proportional to  $n$ . Because of this, just as

in the Brownian example, the value of a firm with  $n$  blueprints is still linear in  $n$ . The Bellman equation (2) holds. The balanced growth conditions are again given by (3)-(5). As before, write  $\mu = G(1, m)$ .

Measured in numbers of blueprints, the possible firm sizes are now  $2^s$ , for all  $s+1 \in \mathbb{N}$ . For a firm with  $n$  blueprints, the variable  $s = \log_2(n)$  measures the number of replications that have occurred since the firm was set up. Along a balanced growth path, the distribution of the number of replications in a cohort of age  $a$  satisfies

$$Dp_{0,a} = -\mu p_{0,a}, \quad Dp_{s+1,a} = \mu p_{s,a} - \mu p_{s+1,a}, \quad s+1 \in \mathbb{N}.$$

Starting from  $p_{1,0} = 1$  this yields

$$p_{s,a} = \frac{1}{s!} e^{-\mu a} (\mu a)^s$$

for all  $s+1 \in \mathbb{N}$ . This is, of course, the Poisson distribution with mean  $\mu a$ . As in all previous examples, there are now two possibilities that depend on whether entrepreneurs create new firms or not. If they do not, and all firms have one blueprint at time  $t = 0$ , then the distribution of log firm size will be Poisson with mean  $\mu t$  at time  $t$ . Alternatively, if  $\mu < \eta$ , then the age distribution of firms has the exponential density  $\eta e^{-\eta a}$ . Combining this with the fact that  $s$  given age  $a$  is Poisson with mean  $\mu a$ , this implies

$$p_s = \int_0^\infty \eta e^{-\eta a} p_{s,a} da = \frac{\eta}{\eta + \mu} \left( \frac{\mu}{\eta + \mu} \right)^s \quad (12)$$

for all  $s+1 \in \mathbb{N}$ . Thus log size is geometrically distributed. The right-tail probabilities for the firm size distribution, with firm size measured in numbers of blueprints, are therefore

$$\sum_{s=\log_2(n)}^\infty p_s = \left( \frac{\mu}{\mu + \eta} \right)^{\log_2(n)} = n^{-\log_2(1+\eta/\mu)}$$

for all  $n$  on the grid  $2^s$ ,  $s+1 \in \mathbb{N}$ . On this grid, these right-tail probabilities converge to  $1/n$  as  $\mu \uparrow \eta$ , just as in the case of the Yule process. Obviously, the sample paths of firm size are extremely unrealistic in this example. But it illustrates, together with the independent replication example, how a thick-tailed size distribution can arise from replication acting at different levels of aggregation within the firm.

### 3.4 The Spin-off Interpretation

When firms can grow forever, the key ingredient in generating a stationary size distribution is entry. In all examples presented above, the amount of entry depends on how

many agents have a comparative advantage in creating blueprints from scratch versus using the replication technology available to incumbent firms. In Chatterjee and Rossi-Hansberg [2007], the amount of entry depends instead on the incentives of an employee to report or sell an idea to his employer, versus keeping the idea and using it to start a new firm. Related, in Franco and Filson [2006] employees can start new firms by copying from their employers.

To show the mechanics, consider the independent replication economy of Section 3.3.1 and suppose that no blueprints can be created from scratch. Rather, among the blueprints created within the firm, a fraction  $1 - \alpha$  become spin-offs. Because there are no independent entrepreneurs, the equilibrium conditions (4) and (5) have to be replaced by  $N(l + m) = 1$  and  $\eta = G(1, m)$ . One can view (3) as determining  $l$  and  $m$  as functions of the wage  $w$ , and then these modified equilibrium conditions determine  $N$  and  $w$ . The rate at which the aggregate number of blueprints grows is still  $\eta$ . But the fact that a fraction  $1 - \alpha$  of all new blueprints are spin-offs means that the rate at which firms accumulate blueprints is  $\mu = \alpha\eta \in (0, \eta)$ . The remaining newly created blueprints are the seeds of new firms, and hence the flow of new firms grows at the rate  $\eta$ . The firm age distribution is again exponential, and hence the size distribution will be (11). This spin-off interpretation is a direct translation of the way Yule [1925, p. 24] constructed his distribution: within a species, every individual replicates at the rate  $\eta$ , and occasionally a mutation occurs that generates a new species.

### 3.5 Heterogeneous Organization Capital

Luttmer [2008] combines independent replication of blueprints with heterogeneity in blueprint quality. In a competitive version of the model, output of a quality- $z$  blueprint is  $zF(1, l)$ . New firms are created with a frontier blueprint quality  $z = Z_t$ , where  $Z_t$  grows at some exogenous rate  $\theta$ . Blueprint quality continues to follow the frontier, but a firm and all its blueprints may experience a one-time reduction in quality at some random time, following an exponentially distributed waiting time with mean  $1/\delta$ . As before, all blueprints can be replicated at rates  $G(1, m)$ , where  $m$  is labor. New blueprints are exact copies of the blueprints from which they were produced. When the firm-wide reduction in quality occurs, incentives to replicate and maintain blueprints are reduced, and this slows down the growth rate of the firm. The Bellman equation (2) requires only minor modification and firms grow at either high or low constant rates along a balanced growth path. If fast-growing firms gain and lose blueprints at rates  $\mu$  and  $\lambda$ , then the tail index of the size distribution is  $\zeta = (\eta + \delta)/(\mu - \lambda)$ , provided that  $\mu > \lambda$  and the distribution



of slow-growing firms does not have an even thicker tail. The combination of rapid initial growth and slower long-term growth allows one to account for the fact that the median age of US firms with 10,000 or more employees is only about 75 years, without the random exit assumption and its anomalous implication for the size of exiting firms. In the resulting account of the data, Gibrat's law does not hold. But an econometrician running short panel regressions of growth rates on size would find that Gibrat's law holds except for small firms, as in Evans [1987] and Hall [1987].

### 3.5.1 Obsolescence

An alternative way to account for the slowdown in firm growth, inspired by the vintage capital model of Hopenhayn [2007], is as follows. Instead of the one-time reduction in quality, suppose that frontier blueprints stop tracking the frontier altogether after an exponentially distributed waiting time with mean  $1/\delta$ .

The Bellman equation (2) now requires a more substantial modification. At time  $t$ , let  $Q_t$  be the value of a blueprint at the frontier, and write  $q_{v,t}$  for the value of a blueprint with a productivity that stopped growing at time  $v \leq t$ . Along a balanced growth path, wages are  $w_t = we^{\theta t}$  and the interest rate is  $r_t = \rho + \theta$ . Conjecture  $Q_t = Qe^{\theta t}$  and  $q_{v,t} = q(t-v)e^{\theta t}$ . The Bellman equation for  $Q$  is then

$$\rho Q = \max_{l,m} \{ZF(1,l) + QG(1,m) - w(l+m)\} + \delta [q(0) - Q], \quad (13)$$

and the function  $q(a)$  must solve

$$\rho q(a) = \max_{l,m} \{Ze^{-\theta a}F(1,l) + q(a)G(1,m) - w(l+m)\} + Dq(a). \quad (14)$$

This implies that frontier blueprints are replicated at a rate  $G(1, M)$ , where  $M$  attains the maximum in (13). Blueprints behind the frontier will be replicated at slower rates, if at all. If the marginal product  $D_2F(1, 0)$  is finite, then there will be a finite age  $A$ , measured since the time a blueprint's productivity stopped growing, at which a blueprint becomes obsolete. For example, if  $F(1, l) = \min\{1, l\}$  then the age of obsolescence is  $A = \ln(Z/w)/\theta$ .

Suppose that  $D_2F(1, 0)$  is indeed finite and that the blueprints that can be created by entrepreneurs are so far behind the frontier as to make them obsolete. Then all new blueprints are produced from the existing stock of blueprints. Along a balanced growth path,  $G(1, M) = \eta + \delta$ , so that the population of frontier blueprints grows at the rate  $\eta$ . Since  $M$  is a function only of wages, this equilibrium condition determines the level

of wages. Given the replication rates implied by (14), it is not difficult to calculate the productivity distribution of blueprints behind the frontier. The equilibrium number of blueprints then follows from clearing the labor market using the employment decision rules implied by (13)-(14).

### 3.5.2 Endogenous Growth

It is easy to turn this into a model of endogenous growth by letting owners of improvable blueprints employ  $u$  units of labor per blueprint to improve the productivity of individual blueprints at a rate  $\theta = R(1, u)$ , where  $R$  is a constant-returns-to-scale production function. Blueprints cease to be improvable at a rate  $\delta$ . As in Boldrin and Levine [2002], a blueprint of a particular productivity level is an input in producing a blueprint of a higher quality level. The technology exhibits constant returns to scale. The incentives to improve blueprints are the same for all blueprints at the frontier that can still be improved, and so all frontier blueprints will be improved at a common rate. This will also be the growth rate of wages and per-capita consumption in the economy. Aggregate growth is driven entirely by the incentives of owners of frontier blueprints to improve the quality of their blueprints.

The spin-off device of Chatterjee and Rossi-Hansberg [2007] can now be used to obtain a firm size distribution with the Pareto-like tail observed in the data. As above, let a firm be a collection of blueprints produced from a common spin-off blueprint. A frontier firm generates spin-offs at the rate  $(1 - \alpha)G(1, M)$ . Suppose all blueprints within a firm stop growing in productivity at the same time. Then employment at firms with frontier blueprints grows at the average rate  $\alpha G(1, M)$ , and the age distribution of firms with frontier blueprints will have a density  $(\eta + \delta)e^{-(\eta + \delta)a}$ . The size distribution of all firms at the frontier is again a Yule distribution, with a tail index  $\zeta = 1/\alpha$ . Since firms that can no longer keep up with the frontier exit in finite time, frontier firms dominate the right tail of the size distribution. In this economy, large firms are firms that have been at the frontier of productivity growth for a long time.

## 3.6 Multiproduct Firms

It has been assumed up to now that every firm produces the same good for a competitive market. Different blueprints describe, for example, different production lines, plants or stores.

Alternatively, Klette and Kortum [2004] consider an economy, based on the quality-ladder model of Grossman and Helpman [1991], in which firm size is a reflection of

the number of different products a firm sells. Recall that in Grossman and Helpman [1991] there is a unit measure of distinct commodities at every point in time. There is a constant population of consumers who have logarithmic preferences. Potential producers in a given market each own a blueprint that describes a linear labor-only technology for producing at a certain level of quality. The producer with the highest quality in a given market takes the whole market by setting a price that reduces the potential profits of the producer with the next highest quality to zero. Entrepreneurs can use labor to create a blueprint with a discrete multiplicative quality improvement over that of the incumbent, at some Poisson rate. Upon the arrival of an improved blueprint, the owner of the new blueprint takes over the market. The combination of an evenly spaced quality ladder (in logs) and logarithmic utility implies that profits in a particular market are independent of quality.

Klette and Kortum [2004] modify this economy by allowing an incumbent in one market to combine  $m$  units of labor with the incumbent's blueprint for that market to generate a blueprint for a randomly selected alternative market, at some Poisson rate  $G(1, m)$ . Just as with blueprints created by entrepreneurs, the quality of the blueprint is a discrete multiplicative improvement over the one being used in that market. A firm is a collection of blueprints for different markets, and firm size can be measured by the number of markets  $n \in \mathbb{N}$  in which it operates. Along a balanced growth path  $\mu = G(1, m)$  for all incumbent producers, and there will be some positive flow  $\nu$  of markets taken over by entrepreneurs. Since there is a unit measure of markets, this implies that incumbents lose markets at a rate  $\lambda = \mu + \nu$ . Let  $p_n$  be the stationary distribution of firm size. Requiring the flows in and out of a state  $n$  to add up to zero gives  $0 = \lambda p_2 - (\mu + \lambda)p_1 + \nu$  and  $0 = \mu(n-1)p_{n-1} + \lambda(n+1)p_{n+1} - (\mu + \lambda)np_n$  for all  $n+1 \in \mathbb{N}$ . This can be written as  $(n+1)p_{n+1} - np_n = (\mu/\lambda)[np_n - (n-1)p_{n-1}]$  and then one can use  $np_n = -\sum_{k=n}^{\infty} [(k+1)p_{k+1} - kp_k]$  to conclude that

$$p_n \propto \frac{1}{n} \left(\frac{\mu}{\lambda}\right)^n.$$

The problem with this distribution is that its right tail is even thinner than that of a geometric distribution. It behaves like  $(\mu/\lambda)^n$  instead of  $n^{-\zeta}$  for some  $\zeta$ . The underlying reason is that firms in this economy cannot grow on average: the number of markets is fixed and positive entry means that the average incumbent has to lose markets.

Random exit can save the day. If firms exit from all their markets at some positive rate  $\delta$ , then  $\lambda = \mu - \delta + \nu$  and hence  $\mu - \lambda = \delta - \nu$  will be positive if  $\delta$  is large enough relative to the rate at which entrepreneurs take over markets. Firms can grow conditional

on survival, and the size distribution will then have a tail index  $\zeta = \delta/(\mu - \lambda)$ . As noted before, a difficulty with this solution is that there will be a lot of exit of large firms. An alternative is to assume that the productivity of a firm’s replication technology is high for new firms and drops to a lower level following some exponentially distributed waiting time with mean  $1/\delta$ .

In the monopolistic competition version of Luttmer [2008], distinct blueprints also describe new products that can be produced using a linear labor-only technology. But no two products in the economy are perfect substitutes. Preferences are as in Dixit and Stiglitz [1977] and the number of goods can grow over time. Because of this, new products created by one firm do not imply the elimination of a product by another firm. When there is population growth at a rate  $\eta$ , the total number of products in the economy will grow at the same rate  $\eta$  along a balanced growth path. Hence positive firm growth and positive entry can go together, resulting in the tail index  $\zeta = \eta/(\mu - \lambda)$  reported above.<sup>8</sup>

### 3.7 Networks, Search and Matching

Replication of organization capital across distinct locations is a natural interpretation of firm growth. In the models considered so far, the only link between different units of the firm comes at the investment stage: existing units can be used to create more units. Removing a particular unit of the firm has no effect on how the others function. This is very different from viewing a large firm as an integrated network. Other parts of a hub-and-spoke airline will be affected when one of its hubs goes down. The model of WalMart in Holmes [2009] has an integrated distribution network at its core. It is quite likely that these types of firms are heavily represented among the largest firms we see in the data.

The accumulation of new “blueprints” need not be the result of firm investment. An important and complementary alternative is product or labor market search. If consumers choose to purchase from suppliers not by randomly sampling firms but by randomly sampling other consumers and following their example (Steindl [1965]), then a firm will gain customers in proportion to how many customers it already has. Similarly, if workers find jobs by contacting other workers (Rees [1966], Granovetter [1974], Burdett and Vishwanath [1988]), then employers will find new employees in proportion to how many workers they already have. This is known as preferential attachment in the

---

<sup>8</sup>Bernard, Redding and Schott [2006] present detailed evidence on the importance of turnover in the products produced by US manufacturing firms.

literature on network formation. Gibrat’s law is a natural outcome and positive entry will produce a Yule-type distribution (Luttmer [2006]). The extensive literature on this topic is presented in Jackson [2008].

### 3.8 Growth Rate Variances and Aggregate Fluctuations

The independent and synchronous replication examples have very different implications for the variance of firm growth. In the case of the Yule process, a firm of size  $n_t$  expands to size  $n_t + 1$  with a probability approximately equal to  $\mu n_t \Delta$  over a small period of time  $\Delta$ . Therefore  $E_t[(n_{t+\Delta} - n_t)/n_t]/\Delta \rightarrow \mu$  and  $\text{var}_t[(n_{t+\Delta} - n_t)/n_t]/\Delta \rightarrow \mu/n_t$  as  $\Delta$  goes to zero. Firm growth rates satisfy Gibrat’s law in a weak sense: only the mean growth rate is independent of size. The realized growth rate of a firm of size  $n$  is, roughly, the sample average of  $n_t$  independent and identically distributed random variables. Such a sample average has a variance that behaves like  $1/n_t$ .<sup>9</sup>

In the case of synchronized replication, a firm of size  $n_t$  expands to size  $2n_t$  with a probability approximately equal to  $\mu \Delta$  over a small period of time  $\Delta$ . Therefore  $E_t[(n_{t+\Delta} - n_t)/n_t]/\Delta \rightarrow \mu$  and  $\text{var}_t[(n_{t+\Delta} - n_t)/n_t]/\Delta \rightarrow \mu$  as  $\Delta$  goes to zero. Since replication events affect the firm as a whole, the variance of the growth rate of the firm is also independent of firm size, just as it is in the Brownian example.

The data seem to be somewhere in between these two extremes. Hymer and Pashigian [1962] long ago noted that the standard deviation of firm growth seems to decline with size but not as fast as  $1/\sqrt{n}$ . More recently, Stanley et al. [1996] report a standard deviation that behaves like  $1/n^{1/6}$  for Compustat employment data. Davis et al. [2006] study growth rate variances over time for both Compustat and Census data. For the year 2000, they report a Compustat sample of only about 8,500 employer firms with an average employment of around 5,300 employees, whereas their Census data set has about 4.7 million firms with an average employment of 18 employees. The respective growth rate standard deviations for 2000 are about .25 and .4 (Davis et al. [2006] Figure 6). In the context of the simple examples given here, the randomness that affects replication must be neither completely dependent nor fully independent across blueprints. More generally, growth rate shocks may occur at different levels of aggregation within the firm. Interesting suggestions for models that can account for this can be found in Stanley et al. [1996] and Sutton [2002].

---

<sup>9</sup>Feller [1951] pointed out that certain branching processes, such as the Yule process, can be approximated for large  $n_t$  by the diffusion  $dn_t = \mu n_t dt + \sigma \sqrt{n_t} dW_t$ , which implies the growth rate variance  $\sigma^2/n_t$ .

As pointed out by Gabaix [2009], the rate at which growth rate variances decline with firm size has implications for aggregate fluctuations when one moves away from the abstraction that there is a continuum of firms. To illustrate, ignore the fact that labor supply is completely inelastic in the formal model presented here. In the example of independent replication across blueprints, firm boundaries do not affect aggregate outcomes, and the rate at which a collection of  $N_t$  blueprints grows has a variance that behaves like  $1/N_t$ . For the whole economy,  $N_t$  will be very large and the aggregate growth rate will have a negligible variance. But in the case of synchronous replication, a firm  $j$  with  $n_{j,t}$  blueprints will have a growth rate variance of  $\text{var}_t [(n_{j,t+\Delta} - n_{j,t})/n_{j,t}] \approx \mu\Delta$ , and hence the number of blueprints in a collection of  $J_t$  independent firms grows at a rate with a variance

$$\text{var}_t \left[ \frac{\sum_{j=1}^{J_t} (n_{j,t+\Delta} - n_{j,t})}{\sum_{j=1}^{J_t} n_{j,t}} \right] \approx \mu\Delta \sum_{j=1}^{J_t} \left( \frac{n_{j,t}}{\sum_{j=1}^{J_t} n_{j,t}} \right)^2.$$

The coefficient multiplying  $\mu\Delta$  is the Herfindahl index for  $\{n_{j,t}\}_{j=1}^{J_t}$ . Given the distribution of log size (12), the mean of  $n = 2^s$  is  $1/(1 - \mu/\eta)$  and this is well defined and finite as long as there is entry. The mean of  $n^2 = 2^{2s}$  is  $1/(1 - 3\mu/\eta)$  if  $\eta/\mu > 3$  and infinite otherwise. This corresponds to a tail index  $\zeta = \log_2(1 + \eta/\mu) > 2$ . Thus for  $\zeta > 2$  the Herfindahl index behaves like  $1/J_t$ , and this will be very small since the number of firms is large. But if  $\zeta \in (1, 2)$ , then the average of  $\{n_{j,t}^2\}_{j=1}^{J_t}$  does not converge as  $J_t$  becomes large and the Herfindahl index will not behave like  $1/J_t$ . Gabaix [2009] shows that the variance of aggregate growth behaves like  $1/J_t^{2(1-1/\zeta)}$  for  $\zeta \in (1, 2]$  and  $1/(2 \ln(J_t))$  in the  $\zeta \downarrow 1$  limit. The data support  $\zeta$  close to 1, and for 6 million firms,  $1/J_t \approx 1.67 \times 10^{-7}$  while  $1/(2 \ln(J_t)) \approx .032$ .<sup>10</sup>

#### 4. HETEROGENEOUS PRODUCTIVITY

Replication is a powerful and natural logic that can explain exponential growth in organization capital, and hence Gibrat's law. A similar logic is far from clear when it comes to productivity growth. As Solow [2005, p.10] notes in the context of aggregate growth, exponential growth "ought to require much more convincing justification than it gets in the standard models of endogenous technological change or accumulation of human capital." Special functional forms such as the Cobb-Douglas production function (as

---

<sup>10</sup>In Durlauf [1993] and Conley and Dapor [2003] aggregate fluctuations are affected by sources of correlation that have nothing to do with firm boundaries.

in Lucas [1978]) or the constant-elasticity-of-substitution utility function also play an unfortunate but central role in deriving Gibrat’s law.<sup>11</sup>

The heterogeneous productivity economies surveyed in this section are based on the Dixit-Stiglitz model of differentiated commodities produced by monopolistic competitors, as in Luttmer [2007]. One advantage of this formulation is that it allows for shifts in tastes across differentiated commodities to play a role in accounting for heterogeneity across firms. But many of the observable implications also apply to models with firms that produce for a competitive output market using a Cobb-Douglas technology with a firm-specific fixed factor. Recent examples are Atkeson and Kehoe [2005] and Luttmer [2009].

As before, flow utility is a logarithmic function of composite consumption, and dynamic consumers discount utility flows at the rate  $\rho$ . The population of consumers in the economy is  $H_t = H e^{\eta t}$  and everyone has one unit of effort per unit of time. The population growth rate  $\eta$  is non-negative.

#### 4.1 Product Market Equilibrium

Every period, there are many differentiated commodities and preferences are as in Dixit and Stiglitz [1977]. The type of a commodity is a state variable  $z$  that can be interpreted as quality. The measure of commodities with a quality level at or below  $z$  is  $A_t[z]$  at time  $t$ . The technology and market structure will be such that all commodities of the same quality  $z$  trade at the same price  $p_t[z]$ . As a result, all commodities of quality level  $z$  are consumed at the same rate, denoted by  $c_t[z]$ . Aggregate consumption of the composite good is

$$C_t = \left( \int (z c_t[z])^{1-1/\gamma} dA_t[z] \right)^{1/(1-1/\gamma)},$$

where  $\gamma > 1$  is the elasticity of substitution. Cost minimization gives the familiar iso-elastic demand curves

$$z c_t[z] = \left( \frac{p_t[z]/z}{P_t} \right)^{-\gamma} C_t,$$

where  $P_t$  is the CES price index  $P_t = \left( \int (p_t[z]/z)^{1-\gamma} dA_t[z] \right)^{1/(1-\gamma)}$ .

All commodities are produced by monopolist producers using the labor-only linear technology  $y_t[z] = l_t[z]$ . Let  $w_t$  be the wage in units of final output. Monopolist produc-

---

<sup>11</sup>Houthakker [1955-1956] shows how the Cobb-Douglas production function can arise from a Leontief technology and Pareto productivities. Anderson, de Palma and Thisse [1992] show how CES utility can be derived from Fréchet taste parameters. Exactly how this might fit together with the size distribution of firms, or perhaps the distribution of consumer wealth, remains an open question.

ers set  $p_t[z]$  equal to  $1/(1 - 1/\gamma)$  times the marginal cost of commodity  $z$ . This yields  $p_t[z] = P_t w_t / (1 - 1/\gamma)$  and the definition of  $P_t$  then implies

$$w_t = (1 - 1/\gamma) N_t^{\frac{1}{\gamma-1}} Q_t, \quad (15)$$

where  $N_t$  is the number of commodities and  $Q_t$  measures average quality,

$$N_t = \int dA_t[z], \quad Q_t = \left( \frac{1}{N_t} \int z^{\gamma-1} dA_t[z] \right)^{\frac{1}{\gamma-1}}.$$

Note that  $Q_t$  is the CES price index applied to prices  $1/z$  and the probability measure  $A_t[z]/N_t$ . Using the demand curves, the constant markup over marginal cost, and the equilibrium real wage, it is easy to see that real revenues for a type- $z$  producer are  $(z/Q_t)^{\gamma-1} C_t / N_t$ . A fraction  $1 - 1/\gamma$  of real revenues goes to labor and the remaining fraction  $1/\gamma$  to profits. Write  $v_t[z]$  for profits of a type- $z$  producer. Then

$$\left( l_t[z], \frac{v_t[z]}{w_t} \right) = \left( 1 - \frac{1}{\gamma}, \frac{1}{\gamma} \right) \left( \frac{z}{Q_t} \right)^{\gamma-1} \frac{C_t}{w_t N_t}. \quad (16)$$

Employment, profits, and revenues all scale with  $(z/Q_t)^{\gamma-1}$ . When commodities are almost perfect substitutes, tiny quality differences lead to large size differences.

Let  $L_t$  denote the aggregate amount of labor used to produce differentiated commodities. Aggregating (16) over all producers gives  $w_t L_t = (1 - 1/\gamma) C_t$  and then (15) implies

$$C_t = N_t^{\frac{1}{\gamma-1}} Q_t L_t. \quad (17)$$

Aggregate consumption increases one for one with average quality and labor, and the elasticity with respect to the number of goods is  $(1/\gamma)/(1 - 1/\gamma)$ .

Everything now depends on how  $A_t[z]$  evolves over time.

## 4.2 One-World Melitz [2003]

The simplest example one can imagine is a closed-economy and continuous-time version of Melitz [2003]. In contrast to the initial example of Section 3, labor and entrepreneurial effort are perfect substitutes in this economy. Anyone can hire  $\lambda_E > 0$  units of labor to generate entry opportunities at a unit Poisson rate. An entry opportunity results in a draw of a quality level  $z$  from a time-invariant distribution of entry qualities. If entry occurs, this quality level will remain constant, until it drops to zero forever, resulting in exit. This happens following an exponentially distributed waiting time. In addition, quality must be “maintained” at a flow cost of  $\lambda_F > 0$  units of labor, or else quality



drops to zero permanently, again resulting in exit. The resulting flow profits  $v_t[z] - w_t\lambda_F$  are determined by (15)-(17).

There is no population growth. Together with the time-invariant entry distribution, this gives rise to a balanced growth path with no growth. Firm numbers, average quality, and wages are constant. As a result, profits are constant and entry opportunities result in actual entry if and only if the quality draw  $z$  from the entry distribution is such that  $v_t[z] - w_t\lambda_F$  is positive. The employment size distribution of firms is completely determined by the quality distribution from which potential entrants draw, truncated at the lowest  $z$  for which profits are non-negative. If the entry productivity distribution is Pareto with a tail index  $\alpha$ , then the employment size distribution of firms will be Pareto with tail index  $\zeta = \alpha/(\gamma - 1)$ . This results in well-defined aggregates if and only if, somehow,  $\alpha > \gamma - 1$ .

In this economy, all size differences are attributed to random draws that happen right before firms enter. In US data, most new firms are very small, and substantial size differences appear only over time. This makes calibration difficult. In particular, estimates of the cost of entry are likely to be biased upward to a significant extent, as entrepreneurs do not anticipate uncertain growth but expect to become an average firm right upon entry. Entry amounts to hitting the jackpot in this economy: given that  $\zeta$  is close to 1, the average firm will be very large.

### 4.3 Stochastic Productivity Growth

Consider the following modification of the Melitz [2003] economy. The technology for generating entry opportunities is the same, but instead of drawing from some time-invariant quality distribution, anyone with an entry opportunity at time  $t$  can start with a common but time-dependent entry quality  $Z_t = Ze^{\theta_E t}$ . Following entry at time  $t$ , quality  $z_{t,a}$  evolves with firm age  $a$  according to

$$d \ln(z_{t,a}) = \theta_I da + \sigma_I dW_a, \quad (18)$$

where  $\{W_a\}_{a \geq 0}$  is a Brownian motion that is independent across firms. For now, both  $\theta_E$  and  $\theta_I$  are taken as parameters. As before, a fixed cost of  $\lambda_F$  units of labor is required to keep the firm alive.

#### 4.3.1 Balanced Growth

We are interested in versions of this economy that have a balanced growth path with a stationary employment size distribution. Conjecture that  $L_t/H_t$  is constant along such a

balanced growth path, so that constant shares of the labor force are used for producing differentiated commodities and for paying entry and fixed costs. The employment size distribution can then only be stationary if  $N_t/H_t$  is constant. Since the labor share  $w_t L_t/C_t$  is constant, it follows that  $C_t/(w_t N_t)$  is constant, and hence a stationary employment size distribution corresponds to a stationary distribution for  $z/Q_t$ , by (16). In particular,  $Z_t/Q_t$  must be constant. Entry and average quality must grow at the same rate  $\theta_E$ . Together with (15), (17), and the fact that the number of firms grows at the rate  $\eta$ , this implies that wages and per-capita consumption grow at a rate  $\kappa$  given by

$$\kappa = \theta_E + \left( \frac{1/\gamma}{1 - 1/\gamma} \right) \eta. \quad (19)$$

That is, growth is driven by a combination of entry productivity growth and gains in variety that result from increases in the number of firms. The rate of incumbent productivity growth  $\theta_I$  plays no role in determining the growth rate of this economy.

It remains to determine if there is indeed a stationary size distribution when the number of firms grows at the rate  $\eta$  and wages and aggregate consumption grow at the rate  $\kappa$  given in (19).

### 4.3.2 Exit and Entry Decisions

Given (16), it is convenient to define the state variable

$$e^{s_t[z]} = \frac{v_t[z]}{\lambda_F w_t} = \frac{1}{\gamma \lambda_F} \left( \frac{z}{Q_t} \right)^{\gamma-1} \frac{C_t}{w_t N_t}. \quad (20)$$

This represents variable profits of a type- $z$  firm at time  $t$ , measured in units of labor, relative to the fixed cost of continuing a firm. For entrants,  $s_t[Z_t] = S$  is a constant. As the firm ages,  $s_t[z_{t-a,a}] = s_a$  evolves with age according to  $ds_a = \mu da + \sigma dW_a$ , where

$$[\mu, \sigma] = (\gamma - 1) [\theta_I - \theta_E, \sigma_I]. \quad (21)$$

Using the tools presented in Dixit and Pindyck [1994], one can verify that the optimal policy is to exit when  $s_a$  falls below some threshold  $B < S$ . The resulting market value of a firm in state  $s$  at time  $t$  is  $w_t \lambda_F V(s)$ , where  $V(s)$  satisfies the Bellman equation

$$\rho V(s) = e^s - 1 + \mu DV(s) + \frac{1}{2} \sigma^2 D^2 V(s)$$

for all  $s \geq B$ , together with two boundary conditions:  $V(B) = 0$  and the requirement that  $V(s)$  is bounded above by a multiple of  $e^s$ . This differential equation can be solved

explicitly. The result is an increasing and convex value function with an asymptote that is linear in  $e^s$  when  $s$  becomes large.

Entry must be such that  $\lambda_E \geq \lambda_F V(S)$ , with equality if entry is positive. Conveniently, the function  $V(\cdot)$  only depends on the rate parameters  $\rho$ ,  $\mu$ , and  $\sigma^2$ . Thus,  $\lambda_E = \lambda_F V(S)$  determines  $S$  if entry is positive.

### 4.3.3 The Stationary Size Distribution

The number of firms in state  $s$  is now  $f(s)N_t$ , where  $f(s)$  is a probability density on  $[B, \infty)$ . It can be shown that the flow of firms that exit at the barrier  $B$  is  $\frac{1}{2}\sigma^2 Df(B)N_t$ . For the number of firms to grow at the rate  $\eta$ , it must be that the entry rate is  $\varepsilon = \eta + \frac{1}{2}\sigma^2 Df(B)$ . The density  $f$  satisfies the Kolmogorov forward equation  $\eta f(s) = -\mu Df(s) + \frac{1}{2}\sigma^2 D^2 f(s)$  for all  $s \in (B, S) \cup (S, \infty)$ . The density is continuous at  $S$ , but it has a kink because of entry. The remaining boundary conditions are  $f(B) = 0$  and the requirements that  $f$  must be positive and integrate to 1 on  $(B, \infty)$ . This results in the stationary density

$$f(s) = \frac{\zeta \zeta_*}{\zeta + \zeta_*} \min \left\{ \frac{e^{(\zeta + \zeta_*)(s-B)} - 1}{e^{\zeta_*(S-B)} - 1}, \frac{e^{(\zeta + \zeta_*)(S-B)} - 1}{e^{\zeta_*(S-B)} - 1} \right\} \times e^{-\zeta(s-B)} \quad (22)$$

where  $\zeta$  and  $\zeta_*$  are defined in (8).

This is essentially the same result as (7), except that the support of (22) has a finite lower bound  $B$ , whereas the  $s = \ln(n)$  version of (7) would have support  $(-\infty, \infty)$ . In fact, the same distribution would arise here if there were no fixed cost. The two distributions are the same conditional on  $s \geq S$ . The difference only shows up below  $S$ , and  $S$  must be small because entering firms are small. Thus differences between (7) and (22) only appear for very small firms. The real payoff of the current formulation is that it has a better prediction about where most of the exit we see in the data should be observed: at the very low end of the distribution.

As in (7) the condition for  $e^s$ , and thus firm employment, to have a finite mean is  $\zeta > 1$ , or equivalently,  $\mu + \frac{1}{2}\sigma^2 < \eta$ . Since firm employment is  $l_t[z] \propto e^{st[z]}$ , this simply says that the mean employment growth rate of incumbent firms not at the exit barrier is less than the population growth rate. Incumbent productivity cannot grow too fast. In contrast to (7), the difficulty is now that there is no equilibrium condition that ensures this will be the case. Furthermore, there is no explanation for why  $\zeta$  is close to 1 in the data.

#### 4.3.4 No Fixed Costs

If there are no fixed costs, a balanced growth path exists even if  $\mu + \frac{1}{2}\sigma^2 \geq \eta$ . In that case, the number of firms is whatever it is at the initial date,  $L_t = H_t$ , and wages and per-capita consumption grow at the rate

$$\theta_I + \frac{1}{2}(\gamma - 1)\sigma_I^2 \geq \theta_E + \left(\frac{1/\gamma}{1 - 1/\gamma}\right)\eta.$$

The left-hand side is just the growth rate of  $Q_t$  when incumbent firms never exit and there is no entry, and the right-hand side is the growth rate that arises when there is non-trivial entry along the balanced growth path. Multiplying both sides by  $\gamma - 1$  and using the definitions (21) of  $\mu$  and  $\sigma^2$  shows that this inequality is just  $\mu + \frac{1}{2}\sigma^2 \geq \eta$ . In this economy, wages are driven up by rapid productivity growth among incumbents, and entry productivity cannot keep up. Aggregate growth is now entirely determined by how fast incumbents grow.

Although there is a balanced growth path, the size distribution is no longer stationary but spreads out forever. If all firms happen to be the same at the initial date, then the distribution of log employment is simply normal with a mean and variance that are linear in time. The employment size distribution of U.S. firms is certainly very different now from what it was a century or two ago, and this type of spreading out may in fact account for what happened during the early part of the twentieth century. But this steady spreading out is hard to reconcile with data for recent decades.<sup>12</sup>

#### 4.3.5 Trade

Firms in the economy just described pay an entry cost once and can then sell to all consumers. One can imagine there are geographically distinct markets, and that entry into each requires a market-specific entry cost. Irarrazabal and Opromolla [2008] do this in a two-country world. In their economy, firms can enter the domestic market at one cost and pay another cost to enter the foreign market. As usual, shipping to the foreign market is also subject to iceberg transportation cost and labor markets are distinct. Firm-specific productivity keeps changing, and as a result some exporters will not be

---

<sup>12</sup>Despite its counterfactual implications, the economy without fixed costs has some distinct advantages in terms of tractability, resulting from the fact that  $Q_t$  suffices to determine aggregate dynamics. The economy is isomorphic to a durable goods economy in which durable goods are produced using labor  $H_t - L_t$  and  $(1 - 1/\gamma)\ln(L_t)$  is added to flow utility. The dynamics can be completely characterized in a simple phase diagram.

selling as much in the domestic market as other firms that never entered the foreign markets. Furthermore, their exports may be small and not enough to recover the cost of setting up the export operation. This is what is seen in the data and could not occur in Melitz [2003].<sup>13</sup>

#### 4.4 The Managerial Productivity Interpretation

Up to now, the quality variable  $z$  has been an attribute of a firm. Anything produced by the firm has quality  $z$ , no matter who is hired to do the work. An alternative is to assume the quality variable  $z$  is an attribute of the manager who runs the firm, as in Lucas [1978]. An entrepreneur can use  $\lambda_E$  units of labor to create an entry opportunity. When the entry opportunity arises, the entrepreneur becomes the manager of the firm. As long as this manager runs the firm, quality evolves according to (18). A manager cannot be supplying labor on the side, and quality drops to zero if the manager decides to quit. The manager's outside opportunity remains to work at the wage  $w_t$ . In this economy,  $\lambda_F = 1$  and the value of being a manager is simply  $w_t(V(s) + 1/\rho)$  when the firm is in state  $s$  at time  $t$ . The manager shuts down the firm and becomes a worker again when this value falls below  $w_t/\rho$ , which is exactly when  $s$  reaches  $B$ .

An obvious difficulty with this very simple story is that we do observe turnover in top management, and most larger firms do not break up when their CEO retires. But the example serves to emphasize that  $V(s)$  need not be the market value of a firm. The quality  $z$  could be due to a manager or team of managers who are in control of the firm. Interesting issues arise when one asks to what extent  $z$  can be transferred from one manager to the next.

One can take this a step further by taking  $z$  to be a composite of firm and managerial attributes. Holmes and Schmitz [1995] take this approach in their analysis of managerial turnover at small businesses. Gabaix and Landier [2008] and Tervio [2008] describe a competitive assignment model, in the tradition of Rosen [1982] and Sattinger [1993], in which there is heterogeneity in both firm and managerial attributes. The quantitative implications of fully dynamic versions of these models remain to be investigated.

#### 4.5 Endogenizing Incumbent Productivity Growth

In the economy described so far, aggregate growth does not depend on how incumbent productivity grows when there is a balanced growth path with a stationary size distrib-

---

<sup>13</sup> Arkolakis [2006] explains the existence of small exporters by introducing a marketing cost that exhibits increasing marginal costs in the number of consumers reached.

ution. But the level of the balanced growth path certainly does. This is one reason to open the black box that produces the  $z_{t,a}$  given in (18).

In a two-country version of the above economy, Atkeson and Burstein [2009] make  $\theta_I$  a choice variable of the firm. Their main assumption is that the cost, in units of labor, of choosing a particular growth rate  $\theta_I$  is equal to  $z^{\gamma-1}$  times an increasing convex function of  $\theta_I$ . Because  $v_t[z]/w_t \propto z^{\gamma-1}$ , the assumption that costs scale with  $z^{\gamma-1}$  is critical to ensure that Gibrat's law holds for large  $z$ . The same proportionality arises in the organization capital examples of Section 3, but the constant-returns-to-scale justification given there is no longer available to motivate the assumption of Atkeson and Burstein [2009].

To provide a motivation for this way of accounting for Gibrat's Law, suppose the fixed cost  $\lambda_F$  includes compensation for a manager who has one unit of effort per unit of time that can be allocated to overseeing current production and improving productivity. Productivity grows at a rate  $\theta_I = g(x)$  when the manager uses  $x$  units of effort to improve productivity. Current productivity is  $z \times (h(y))^{1/(\gamma-1)}$  instead of  $z$  if the manager uses  $y$  units of effort to manage current production. So managerial attention has a multiplicative effect on current quality. Because of the unique position the manager is in, neither task can be delegated or outsourced, and so  $x$  and  $y$  have to add up to the one unit of effort available to the manager. Both  $g$  and  $h$  are increasing and smooth, and  $h$  is positive. The Bellman equation for the value of the firm is then

$$\rho V(s) = \max_{x \in [0,1]} \left\{ e^s h(1-x) - 1 + (\gamma-1)(g(x) - \theta_E) DV(s) + \frac{1}{2} \sigma^2 D^2 V(s) \right\},$$

and  $V(B) = 0$  at an exit boundary  $B$ . Conjecture that  $x$  is approximately independent of  $s$  for large  $s$ . Recall that the present value of  $e^s$  received in perpetuity is  $e^s / (\rho - [\mu + \sigma^2/2])$  if the drift of  $s$  is a constant  $\mu$ . For large  $s$ , the fixed cost has a minor effect on the value of the firm, and thus one expects  $V(s)$  to behave like  $e^s$  when  $s$  is large. In turn, the optimal allocation of managerial effort satisfies  $e^s Dh(1-x) = (\gamma-1) Dg(x) DV(s)$  if it is interior, and this then confirms the conjecture. In words, an approximate version of Gibrat's law arises here because the tasks of overseeing production and improving productivity must be assigned to the same manager.

#### 4.6 Endogenizing Entry Productivity Growth

The entry quality process  $Z_t$  has up to now been viewed as an independent source of ideas that drives aggregate growth when it increases fast enough relative to what incumbent producers can come up with. Perhaps some fraction of the population is motivated

simply by the bragging rights of having raised  $Z_t$  to  $Z_t(1 + \theta_E dt)$ , for the public good. Perhaps the economy has a sector that replicates and improves blueprints along the lines of Section 3.5, and entrepreneurs need these blueprints to start new firms.

An alternative is to assume that ideas are generated by incumbent firms and that  $Z_t$  is the result of imitation by entrants. This type of spillover goes back to Arrow [1962] and has played a dominant role in modeling endogenous growth over the past two decades. Take  $\theta_I$  again to be exogenous and interpret the quality process (18) as the outcome of learning by doing or experimentation by incumbent firms. Firms choose to stop this process and exit when their quality reaches a level  $X_t < Z_t$  that solves  $s_t[X_t] = B$ . Entrants imitate and improve the quality of exiting firms. The entry cost  $\lambda_E$  allows entrants to start with a quality  $e^{(\gamma-1)\Delta} X_t$  for some  $\Delta > 0$ . From (20), this implies  $S - B = \Delta$ . Along a balanced growth path with positive entry, the cost of entry must match the value  $V(S)$  of newly created firms. Hence

$$\lambda_E = \lambda_F V(B + \Delta). \quad (23)$$

Since  $B$  and  $V(\cdot)$  depend only on  $\rho$ ,  $\mu = (\gamma - 1)(\theta_E - \theta_I)$  and  $\sigma = (\gamma - 1)\sigma_I$ , this equilibrium condition determines  $\theta_E$ .

The simple equilibrium condition (23) leads to straightforward comparative statics. An increase in the equilibrium value of  $\theta_E$  lowers the function  $V(\cdot)$ , as incumbents fall behind more quickly. It follows that  $\theta_E$  is increasing in  $\Delta$ , as the higher value of entry implied by an increase in  $\Delta$  is off-set in equilibrium by a higher growth rate  $\theta_E$ . Greater improvements by entrants raise the growth rate of the economy. More rapid increases in incumbent quality also raise the growth rate of the economy: an increase in the incumbent growth rate  $\theta_I$  leads to a one-for-one increase in  $\theta_E$ , so that  $\mu$  and therefore  $V(\cdot)$  remain unchanged. In sharp contrast to what happens when entry productivity is an independent source of ideas, the equilibrium growth rate of the economy now responds one for one to changes in incumbent growth.

#### 4.6.1 Ensuring Stationarity

A difficulty is, again, that there may not be a balanced growth path with a stationary size distribution. The equilibrium value of  $\mu$  implied by (23) does not guarantee  $\eta > \mu + \sigma^2/2$ , and if this condition is violated then the aggregate demand for labor by firms is infinite.

This turns out to be an artifact of the assumption, maintained so far in Section 4, that entrepreneurial effort and labor are perfect substitutes. This makes the equilibrium condition (23) depend only on growth rates, and not on the number of firms. If instead

the supplies of labor and entrepreneurial effort are not perfectly elastic, as in Section 3, then the equilibrium conditions become

$$\begin{bmatrix} E(V(B + \Delta)) \\ L(V(B + \Delta)) \end{bmatrix} = \begin{bmatrix} \eta + \frac{1}{2}\sigma^2 Df(B) \\ \lambda_F (1 + (\gamma - 1) \int_B^\infty e^s f(s) ds) \end{bmatrix} \frac{N}{H}, \quad (24)$$

where  $E(\cdot)$  and  $L(\cdot)$  are the supplies of entrepreneurial effort and labor introduced in Section 3.1. The first condition ensures that there is enough entry to make the number of firms grow at the population growth rate. The second is the labor market clearing condition. The value function  $V$ , the exit barrier  $B$  and the stationary density  $f$  are all functions of  $\theta_E$ , via  $\mu$ . These two conditions therefore jointly determine  $\theta_E$  and  $N/H$ , and the labor market clearing condition forces the mean of  $e^s$  to be finite (Luttmer [2009].)

#### 4.7 Alternative Entry and Fixed Cost Assumptions

The assumption that some type of labor is essential for starting and continuing firms is a key ingredient in generating a stationary employment size distribution. The following two examples illustrate what can happen when one deviates from this assumption.<sup>14</sup>

##### 4.7.1 Romer [1990]

Famously, Romer [1990] has an important scale effect that arises from the assumption that the cost of introducing new goods, measured in units of labor, declines with the existing number of goods. Consider, therefore, what happens when there is no population growth. The entry cost in Romer [1990] is  $\lambda_E/N_t$  units of labor, reflecting a positive externality from having many producers. Balanced growth can arise only if the fixed cost is taken to be  $\lambda_F/N_t$ . The growth rate of consumption and wages is  $\kappa = \theta_E + \omega/(\gamma - 1)$  instead of (19), where  $\omega$  is the equilibrium growth rate of the number of commodities. Along a balanced growth path  $C_t$  and  $w_t$  grow at the same rate, and so  $e^{st[z]}$ , as defined in (20) with  $\lambda_E$  replaced by  $\lambda_E/N_t$ , continues to be proportional to  $(z/Q_t)^{\gamma-1}$ . Since both entry and fixed costs are proportional to  $1/N_t$ , the equilibrium condition with positive entry is  $\lambda_E = \lambda_F V(S)$ , and this determines  $S$  as before. The Kolmogorov forward equation for the size distribution is now  $\omega f(s) = -\mu Df(s) + \frac{1}{2}\sigma^2 D^2 f(s)$ , and this implies a size distribution (22), but with  $\zeta$  and  $\zeta_*$  now a function of  $\omega$  instead of  $\eta$ . The remaining condition for a balanced growth path is now the labor market clearing

---

<sup>14</sup>An appendix with detailed calculations for this section can be found at [www.luttmer.org](http://www.luttmer.org).



condition

$$H = \lambda_E \left( \omega + \frac{1}{2} \sigma^2 Df(B) \right) + \lambda_F \left( 1 + (\gamma - 1) \int_B^\infty e^s f(s) ds \right), \quad (25)$$

which no longer depends on  $N$  because entry and fixed costs are  $(\lambda_F, \lambda_E)/N_t$ . This equilibrium condition determines the growth rate  $\omega$ , and the scale effect is apparent from the fact that this equilibrium condition depends on  $H$ . By construction, (25) will force  $\omega$  to be such that the mean of  $e^s$  is finite. As in (24), the fact that the growth rate of the economy is jointly determined with the labor market clearing condition ensures that the stationary size distribution will have a finite mean. But it is important to note here what is actually stationary: the state variable  $s_t[z_{t-a,a}]$ . The definition (20) with  $\lambda_F$  replaced by  $\lambda_F/N_t$  tells us that  $z_{t-a,a}/Q_t$  will be stationary, and then (16) implies that  $N_t l_t[z_{t-a,a}]$  is stationary. Since  $N_t$  grows at the rate  $\omega$ , the employment size distribution cannot be stationary. As the number of firms and goods in the economy grows over time, while the population is constant, the average firm must shrink at an exponential rate as the economy moves along its balanced growth path. This does not fit the data.

#### 4.7.2 Atkeson and Burstein [2009]

Atkeson and Burstein [2009] take entry and fixed costs to be proportional to a Cobb-Douglas composite of labor and final output. To describe what this does, consider the extreme case in which entry and fixed costs are constant in units of final output. This is the lab-equipment specification of Rivera-Batiz and Romer [1991]. Interpret  $\lambda_E$  and  $\lambda_F$  to be the cost parameters in units of final output and define  $e^{s_t[z]} = v_t[z]/\lambda_E$  instead of  $e^{s_t[z]} = v_t[z]/(\lambda_F w_t)$ . Conjecture that there is a balanced growth path with the number of firms growing at some rate  $\omega$ . With firms exiting at a constant exit barrier  $B$ , this implies an entry rate  $\varepsilon = \omega + \frac{1}{2} \sigma^2 Df(B)$ . The aggregate resource constraint is then

$$C_t + (\lambda_E \varepsilon + \lambda_F) N_t = N_t^{\frac{1}{\gamma-1}} Q_t H_t,$$

instead of (17). Balanced growth requires that  $C_t$  and  $N_t$  grow at the common rate  $\omega$ , and then the right-hand side of this resource constraint implies  $\omega = \theta_E + \eta + \omega/(\gamma - 1)$ . Provided  $\gamma \neq 2$ , this yields

$$\omega = \frac{\theta_E + \eta}{1 - \frac{1/\gamma}{1-1/\gamma}}.$$

It is easy to see that the calculations in Sections 4.3.2 and 4.3.3 continue to apply, with  $\eta$  replaced by  $\omega$ . In particular, the modified definition of  $s_t[z]$  implies that the distribution

of  $z_{t-a,a}/Q_t$  is stationary, and the mean of  $e^{s_t[z_t-a,a]}$  will be finite if  $\mu + \frac{1}{2}\sigma^2 < \omega$ . Per-capita consumption and wages grow at the rate  $\kappa = \omega - \eta$ . That is, growth manifests itself entirely in the difference between the growth rate of the number of firms and the population growth rate. If  $\gamma > 2$ , this is positive if  $\theta_E + \eta$  is positive, and thus average firm employment must decrease exponentially over time. The firm employment size distribution cannot be stationary. In Atkeson and Burstein [2009], these complications do not come to the fore because  $\theta_E$  and  $\eta$  are both assumed to be zero, and hence  $\omega = 0$  if  $\gamma > 2$ . In the knife-edge case of  $\gamma = 2$ , aggregate output simplifies to  $N_t Q_t H_t$  and there is a balanced growth path if  $\theta_E = \eta = 0$ . The resulting economy is an AK economy and if entry costs are not too high, the number of firms will grow at some positive rate  $\omega$ . The employment size distribution must again be non-stationary.

## 5. CONCLUDING REMARKS

Firms can be large for many reasons. They can be highly productive and face only weak decreasing returns to scale, or produce a variety that is a close substitute for what others can produce. They can be marginally more productive than other firms and have had enough time to grow and exploit this advantage. Their managers may be highly skilled and therefore control larger teams of employees or larger hierarchies. A better account of the quantitative importance of each of these interpretations is needed. Such an account can tell us more about what drives aggregate productivity growth. More generally, it can serve as a backbone for attempts to improve our understanding of the aggregate economy. Why does unemployment take such a long time to come down following a recession? Perhaps some of the work discussed here can provide a basis for better models of how new and existing firms create jobs.

A common thread in the models surveyed in this paper is the importance of selection and reallocation as a mechanism for getting the economy to grow. Blueprints that can no longer be improved fall behind and are discarded, while frontier blueprints are improved and replicated rapidly. Firms that achieve few productivity improvements are forced to exit and are replaced by entrants that can do better. Selection can operate at many levels, ranging from the level of individual employees to that of firms as a whole. Trial and error combined with selection is a mechanism that can be helped by foresight but does not require it, and the large gross flows we see at all levels indicates that getting productivity improvements right is difficult. The empirical work of Restuccia and Rogerson [2008] and Hsieh and Klenow [2009] suggests that impediments to this reallocation process can have large effects on aggregate output.

## REFERENCES

- [1] Acemoglu, D., *Introduction to Modern Economic Growth*, Princeton University Press (2008).
- [2] Aghion, P. and S.N. Durlauf, *Handbook of Economic Growth*, North Holland (2005).
- [3] Alvarez, F.E., F.J. Buera, and R.E. Lucas, Jr., “Models of Idea Flows,” NBER working paper 14135 (2008).
- [4] Anderson, S.P., A. de Palma and J.-F. Thisse, *Discrete Choice Theory of Product Differentiation*, MIT Press (1992).
- [5] Arkolakis, C., “Market Access Costs and the New Consumers Margin in International Trade,” working paper, University of Minnesota (2006).
- [6] Arkolakis, C., “A Unified Theory of Firm Selection and Growth,” working paper, Yale University (2009).
- [7] Arrow, K.J. “The Economic Implications of Learning by Doing,” *Review of Economic Studies*, vol. 29, no. 3, 155-173 (1962).
- [8] Atkeson, A. and A. Burstein, “Innovation, Firm Dynamics, and International Trade,” working paper, UCLA (2009).
- [9] Atkeson, A., and P. Kehoe, “Modelling and Measuring Organization Capital,” *Journal of Political Economy*, vol. 113, no. 5, 1026-1051 (2005).
- [10] Benhabib, J. and A. Bisin, “The Distribution of Wealth and Redistributive Policies” New York University (2006).
- [11] Bernard, A.B., J. Eaton, J.B. Jensen, and S. Kortum, “Plants and Productivity in International Trade,” *American Economic Review*, vol. 93, no. 4, 1268-1290 (2003).
- [12] Bernard, A.B., S.J. Redding, and P.K. Schott, “Multi-Product Firms and Product Switching,” NBER working paper no. 12293 (2006).
- [13] Boldrin, M. and D.K. Levine, “Factor Saving Innovation,” *Journal of Economic Theory*, vol. 105, 18-41 (2002).
- [14] Burdett, K. and T. Vishwanath, “Balanced Matching and Labor Market Equilibrium,” *Journal of Political Economy*, vol. 96, no. 5, 1048-1065 (1988).

- [15] Chatterjee, S. and E. Rossi-Hansberg, “Spin-Offs and the Market for Ideas,” working paper, NBER working paper 13198 (2007).
- [16] Conley, T.G. and B. Dupor, “A Spatial Analysis of Sectoral Complementarity,” *Journal of Political Economy*, vol. 111, no. 2, 311-352, (2003).
- [17] Davis, S.J., J. Haltiwanger, R. Jarmin, and J. Miranda, “Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms,” *NBER Macroeconomics Annual 2006, Volume 21*, edited by D. Acemoglu, K. Rogoff and M. Woodford, 107-156.
- [18] Dixit, A.K., and J.E. Stiglitz, “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review*, LXVII, 297-308 (1977).
- [19] Dixit, A.K. and R.S. Pindyck, *Investment under Uncertainty*, Princeton University Press (1994).
- [20] Dunne, T., M.J. Roberts, and L. Samuelson, “The Growth and Failure of U.S. Manufacturing Plants,” *Quarterly Journal of Economics*, CIV, 671-698 (1989).
- [21] Durlauf, S.N. “Nonergodic Economic Growth,” *The Review of Economic Studies*, vol. 60, no. 2, 349-366 (1993).
- [22] Eaton, J. and S. Kortum, “International Technology Diffusion: Theory and Measurement,” *International Economic Review*, vol. 40, no. 3, 537-570 (1999).
- [23] Eaton, J., S. Kortum and F. Kramarz, “An Anatomy of International Trade: Evidence From French Firms,” NBER working paper 14610 (2008).
- [24] Ericson, R. and A. Pakes, “Markov-Perfect Industry Dynamics: A Framework for Empirical Work,” *Review of Economic Studies*, vol. 62, no. 1, 53-82 (1995).
- [25] Evans, D.S., “The Relationship between Firm Growth, Size and Age: Estimates for 100 Manufacturing Industries,” *Journal of Industrial Economics*, vol. XXXV, no. 4, 567-581 (1987).
- [26] Feller, W., “Diffusion Processes in Genetics,” *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Edited by J. Neyman, University of California Press, Berkeley, CA (1951), 227-246.

- [27] Feller, W., *An Introduction to Probability Theory and Its Applications, Volume I, Third Edition*, Wiley (1968).
- [28] Franco, A. and D. Filson, “Spin-Outs: Knowledge Diffusion through Employee Mobility,” *RAND Journal of Economics*, vol. 37, no. 4, 841-860 (2006).
- [29] Gabaix, X., “Zipf’s Law for Cities: An Explanation,” *Quarterly Journal of Economics*, vol. 114, no. 3, 739-767 (1999).
- [30] Gabaix, X. and A. Landier, “Why Has CEO Pay Increased so Much,” *Quarterly Journal of Economics*, vol. 123, no. 1, 49-100 (2008).
- [31] Gabaix, X. “Power Laws in Economics,” *Annual Reviews in Economics*, vol. 1, 255-294 (2009).
- [32] Gibrat, R., *Les Inégalités économiques*, Paris, France: Librairie du Recueil Sirey (1931).
- [33] Granovetter, M. *Getting a Job*, University of Chicago Press (1974).
- [34] Grossman, G.M. and E. Helpman, *Innovation and Growth in the Global Economy*, MIT Press (1991).
- [35] Hall, B. H., “The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector,” *Journal of Industrial Economics*, vol. XXXV, no. 4, 583-605 (1987).
- [36] Hayashi, F., “Tobin’s Marginal q and Average q: A Neoclassical Interpretation,” *Econometrica*, vol. 50, no. 1, 213-224 (1982).
- [37] Helpman, E., M.J. Melitz and S.R. Yeaple, “Export Versus FDI with Heterogeneous Firms,” *American Economic Review*, vol. 94, no. 1, 300-316 (2004).
- [38] Holmes, T.J., “The Diffusion of Wal-Mart and Economies of Density,” working paper, University of Minnesota (2009).
- [39] Holmes, T.J. and J.A. Schmitz, Jr., “On the Turnover of Business Firms and Business Managers,” *Journal of Political Economy*, vol. 103, no. 5, 1005-1038 (1995).
- [40] Hopenhayn, H., “Entry, Exit and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, vol. 60, no. 5, 1127-1150 (1992).

- [41] Hopenhayn, H. “Knowledge, Diffusion and Reallocation,” UCLA (2007).
- [42] Houthakker, H.S. “The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis,” *The Review of Economic Studies*, vol. 23, no. 1, 27-31 (1955-1956).
- [43] Hsieh, C.-T. and Klenow, P.J., “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, vol. 124, no. 4, 1403-1448 (2009).
- [44] Hymer, S. and P. Pashigian, “Firm Size and Rate of Growth,” *Journal of Political Economy*, vol. 70, no. 6, 556-569 (1962).
- [45] Ijiri, Y., and H.A. Simon, “Business Firm Growth and Size,” *American Economic Review*, LIV, 77-89 (1964).
- [46] Irarrazabal, A.A. and L.D. Opromolla, “A Theory of Entry and Exit into Exports Markets,” working paper, Banco de Portugal (2008).
- [47] Jackson, M.O., *Social and Economic Networks*, Princeton University Press (2008).
- [48] Karlin, S. and H.M. Taylor, *A First Course in Stochastic Processes*, Academic Press (1975).
- [49] Kendall, D.G., “On the Generalized ”Birth-and-Death” Process,” *The Annals of Mathematical Statistics*, vol. 19, no. 1, 1-15 (1948).
- [50] Klette, T.J., and S. Kortum, “Innovating Firms and Aggregate Innovation,” *Journal of Political Economy*, vol. 112, no. 5, 986-1018 (2004).
- [51] Kortum, S., “Research, Patenting, and Technological Change,” *Econometrica*, vol. 65, no. 6, 1389-1419 (1997).
- [52] Lucas, R.E., Jr., “Adjustment Costs and the Theory of Supply,” *Journal of Political Economy*, vol. 75, no. 4, part 1. 321-334 (1967).
- [53] Lucas, R.E., Jr. “On the Size Distribution of Business Firms,” *Bell Journal of Economics*, vol. 9, no. 2, 508-523 (1978).
- [54] Luttmer, E.G.J., “Consumer Search and Firm Growth,” Federal Reserve Bank of Minneapolis working paper no. 645 (2006).

- [55] Luttmer, E.G.J., "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, vol. 122, no. 3, 1103-1144 (2007).
- [56] Luttmer, E.G.J., "On the Mechanics of Firm Growth," Federal Reserve Bank of Minneapolis working paper 657 (2008).
- [57] Luttmer, E.G.J., "Technology Diffusion and Growth," Federal Reserve Bank of Minneapolis working paper 672 (2009).
- [58] Neal, D. and S. Rosen, "Theories of the Distribution of Earnings," in *Handbook of Income Distribution*, vol. 1, edited by A.B. Atkinson and F. Bourguignon, Elsevier (2000), Chapter 7, 379-427.
- [59] McGrattan, E.R. and E.C. Prescott, "Unmeasured Investment and the Puzzling U.S. Boom in the 1990s," Federal Reserve Bank of Minneapolis, Research Staff Report 369 (2009).
- [60] Melitz, M.J. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, vol. 71, no. 6, 1695-1725 (2003).
- [61] Mitzenmacher, M.A., "Brief History of Generative Models for Power Law and Log-normal Distributions," *Internet Mathematics*, vol. 1, no. 2: 226-251 (2004).
- [62] Prescott, E.C. and M. Visscher, "Organization Capital," *Journal of Political Economy*, vol. 88, no. 3, 446-461 (1980).
- [63] Reed, W.J., "The Pareto, Zipf and other power laws," *Economics Letters*, vol. 74, 15-19 (2001).
- [64] Rees, A., "Information Networks in Labor Markets," *American Economic Review*, vol. 56, no. 1/2, 559-566 (1966).
- [65] Restuccia, D. and R. Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, vol. 11, 707-720 (2008).
- [66] Rivera-Batiz, L.A. and P. M. Romer, "Economic Integration and Endogenous Growth," *Quarterly Journal of Economics*, vol. 106, no. 2, 531-555 (1991).
- [67] Romer, P.M., "Endogenous Technical Change," *Journal of Political Economy*, vol. 98, no. 5, part 2, S71-S102 (1990).

- [68] Rosen, S., “Authority, Control, and the Distribution of Earnings,” *Bell Journal of Economics*, vol. 13, no. 2, 311-323 (1982).
- [69] Ross, S.M., *Stochastic Processes, Second Edition*, Wiley (1996).
- [70] Rossi-Hansberg, E. and M.L.J. Wright, “Establishment Size Dynamics in the Aggregate Economy,” *American Economic Review*, vol. 97, no. 5, 1639-1666 (2007).
- [71] Roy, A.D., “Some Thoughts on the Distribution of Earnings,” *Oxford Economic Papers*, New Series, vol. 3, no. 2, 135-146 (1951).
- [72] Sattinger, M., “Assignment Models of the Distribution of Earnings,” *Journal of Economic Literature*, vol. 31, no. 2, 831-880 (1993).
- [73] Simon, Herbert A., “On a Class of Skew Distribution Functions,” *Biometrika*, vol. 42, no. 3/4, 425-440 (1955).
- [74] Solow, R.M., “Reflections on Growth Theory,” in *Handbook of Economic Growth*, vol. 1A, edited by Aghion, P. and S.N. Durlauf, North Holland (2005), 3-10.
- [75] Stanley, R. Michael, L.A.N. Amaral, S.V. Buldyrev, S. Harlin, H. Leschorn, P. Maass, M.A. Salinger, H.E. Stanley, Scaling Behavior in the Growth of Companies, *Nature* 319, 804-806 (1996).
- [76] Steindl, J., *Random Processes and the Growth of Firms; A Study of the Pareto Law*, Hafner Publishing Company, New York, NY (1965).
- [77] Sutton, J., “Gibrat’s Legacy,” *Journal of Economic Literature*, vol. XXXV, 40-59 (1997).
- [78] Sutton, J., “The Variance of Firm Growth Rates: The ‘Scaling’ Puzzle,” *Physica A*, CCCXII, 577-590 (2002).
- [79] Terviö, M., “The Difference That CEOs Make: An Assignment Model Approach,” *American Economic Review*, vol. 98, no. 3, 642-668 (2008).
- [80] Yule, G.U., “A Mathematical Theory of Evolution Based on the Conclusions of Dr. J.C. Willis, F.R.S.,” *Philosophical Transactions of the Royal Society of London, Series B, Containing Papers of a Biological Character*, vol. 213, 21-87 (1925).
- [81] Weintraub, G.Y., C.L. Benkard, and B. Van Roy, “Markov Perfect Industry Dynamics with Many Firms,” *Econometrica*, vol. 76, no. 6, 1375-1411 (2008).