

On the Mechanics of Firm Growth

ERZO G. J. LUTTMER

*University of Minnesota and Federal Reserve Bank of
Minneapolis*

First version received October 2008; final version accepted June 2010 (Eds.)

The Pareto-like tail of the size distribution of firms can arise from random growth of productivity or stochastic accumulation of capital. If the shocks that give rise to firm growth are perfectly correlated within a firm, then the growth rates of small and large firms are equally volatile, contrary to what is found in the data. If firm growth is the result of many independent shocks within a firm, it can take hundreds of years for a few large firms to emerge. This paper describes an economy with both types of shocks that can account for the thick-tailed firm size distribution, high entry and exit rates, and the relatively young age of large firms. The economy is one in which aggregate growth is driven by the creation of new products by both new and incumbent firms. Some new firms have better ideas than others and choose to implement those ideas at a more rapid pace. Eventually, such firms slow down when the quality of their ideas reverts to the mean. As in the data, average growth rates in a cross section of firms will appear to be independent of firm size, for all but the smallest firms.

Key words: Firm size distribution, Gibrat's law, Aggregate growth

JEL Codes: L1, O4

1. INTRODUCTION

Why does the employment size distribution of U.S. firms look like a Pareto distribution, with the fraction of firms with more than n employees roughly equal to $n^{-\zeta}$? Why is the tail index $\zeta \approx 1.05$ barely high enough for the distribution to have a finite mean? More than half of all firms with any employees have no more than four employees. But there are also almost a 1000 firms with more than 10000 employees each, and these firms employ as much as a quarter of the U.S. labour force. What accounts for the large amount of heterogeneity in firm size? How does this heterogeneity evolve over time? Some benchmark answers to these questions are needed for the systematic use of firm-level data in the study of aggregate growth and fluctuations.

In the presence of decreasing returns or downward sloping firm demand curves, it is possible that the highly skewed size distribution entirely reflects a highly skewed productivity distribution. Such a productivity distribution can arise if productivity growth is random and only sufficiently productive firms can survive. Given isoelastic cost functions or demand curves, random productivity growth gives rise to Gibrat's law, which holds that firm growth rates are independent of size. A stationary size distribution results if employment at incumbent firms grows more slowly on average than aggregate employment. This distribution has a tail index ζ just above 1

if cost parameters are such that there is only a small gap between entrant and incumbent mean productivity growth rates (Luttmer, 2007).¹

This paper associates firm size not primarily with productivity differences, but with organization capital (Prescott and Visscher, 1980) that can be accumulated through investment over time. In the model, a firm produces one or more differentiated commodities using labour and commodity-specific blueprints. An entrepreneur can set up a new firm by producing a start-up blueprint. After that, the firm can use labour and any of its blueprints to attempt to produce more blueprints for new commodities. Individual blueprints can also become obsolete. The arrival rates of these two types of events are independent and independent across blueprints. Absent other sources of heterogeneity, this implies that the mean growth rate of a firm with more than a single blueprint is independent of firm size—a weak version of Gibrat’s law. Averaging within the firm implies that the variance of firm growth is inversely proportional to firm size, a violation of the strong form of Gibrat’s law according to which the entire distribution of growth rates is independent of firm size. The economy exhibits balanced growth, and increases in variety add to the aggregate growth rate, as in Romer (1990) and Young (1998). As long as there is entry, the size distribution will be stationary with a right tail that behaves like $n^{-\zeta}$.

Independent within-firm replication avoids a problem that arises in economies with only firm-wide productivity shocks. In Luttmer (2007), it takes a standard deviation of firm employment growth of about 40% per annum to jointly account for the size distribution and the 11% rate of firm entry observed in the data. This standard deviation is within the range reported by Davis *et al.* (2007) for all firms, but implausibly high for large firms. Here, large firms are very stable even when small-firm growth rates are sufficiently volatile to be consistent with the observed entry and exit rates. In the simplest version of the model, though, this is too much of a good thing and leads to a rather dramatic counterfactual implication: the median age of firms with more than 10 000 employees is implied to be about 750 years. Stationarity and the weak version of Gibrat’s law force mean incumbent growth rates to be below the growth rate of the aggregate labour force, only about 1% per annum, and averaging within the firm reduces variance by too much for “lucky” firms to become large in a relatively short amount of time.

Newly collected data show that the median age of firms with more than 10 000 employees in 2008 was only about 75 years. With a 40% standard deviation of employment growth, an economy like Luttmer (2007) predicts about 100 years.² But to account for the relatively young age of large firms observed in the data, without assuming there is a 30% chance that employment at WalMart will grow or shrink by more than 40% over the next year, requires abandoning Gibrat’s law.

Suppose therefore that some new firms enter with an initial blueprint of a higher quality than other blueprints in the economy. The resulting higher profits per blueprint create an incentive to copy these blueprints at a higher rate if quality is inherited. If copies stay within the firm, then these new firms will grow fast. If a firm’s quality advantage is transitory, this rapid growth will come to an end eventually. A stationary distribution with a tail index ζ above 1 results if there is positive entry along the balanced growth path. A simple formula shows that this tail index will be close to 1 if firms with high-quality blueprints grow at an equilibrium rate that is

1. The $\zeta = 1$ asymptote is known as Zipf’s law. See Axtell (2001) for recent evidence on the firm size distribution showing that ζ slightly above 1 fits the data well. Well-known empirical studies on Gibrat’s law for firms, based on growth rate regressions that correct for selection, are Evans (1987) and Hall (1987). Sutton (1997) surveys the literature. Gabaix (1999) uses Gibrat’s law to interpret the city size distribution and contains many useful references on the history of the subject. Rossi-Hansberg and Wright (2007) develop a model of the firm size distribution in which there are many industries and the firm size in any given industry follows a stationary process instead of the non-stationary process implied by Gibrat.

2. A new calibration is available at www.luttmer.org.

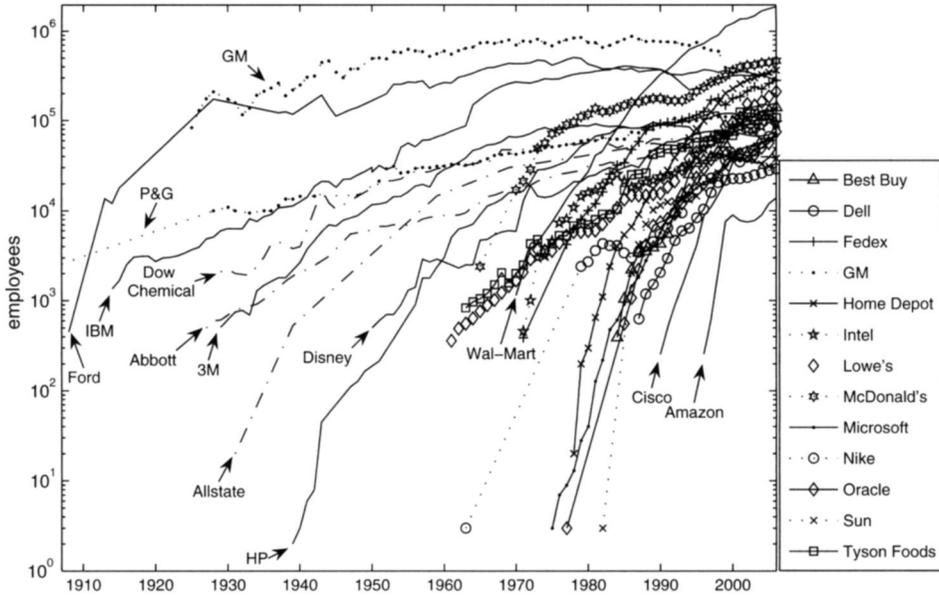


FIGURE 1
Selected growth histories

slightly below the sum of the growth rate of the aggregate labour force and the hazard rate with which high-quality firms lose their edge. Thus, high-quality firms can grow fast if the period of rapid growth is not expected to last too long. But there will be variation in how long firms are in this rapid growth phase, and this variation allows for the appearance of young large firms. This version of the organization capital interpretation of firm growth can match the overall size distribution, the amount of entry and exit, as well as the relatively young age of large firms. Furthermore, although Gibrat's law does not hold, the mean growth rates of surviving firms behave like they do in the data: roughly independent of size for most firms and significantly higher for the smallest firms (Dunne, Roberts and Samuelson, 1989).

Figure 1 presents some corroborating evidence for the type of histories of firm growth predicted by the model. It shows the employment histories of 25 of the nearly 1000 large firms that had more than 10 000 employees in 2008 (the data are described in Appendix A). The average employment growth rate across all firms reported in Figure 1 is almost 18% per annum, and there is considerable variation. In particular, firm growth rates seem to be much above average when firms are relatively small and decline significantly when firms become large. The data shown in Figure 1 represent only a small sample from a population of slightly under a thousand large firms. In turn, this population of large firms was selected over many years from the population of all firms that were ever set up. In U.S. data, the number of firms grows at an annual rate about equal to the 1% growth rate of aggregate employment. Combined with an entry rate of 11%, a steady-state calculation implies that the number of firms that was ever set up is roughly 11 times the current population of around 6 million firms.³ The thousand or so firms with ten thousand or more employees are thus a highly selected sample from a universe of about 66 million firms. In such a selected sample, one might conjecture, it is not surprising to see that large firms tend to

3. The growth rate of the collection of all historical firms equals the entry rate times the fraction of all historical firms that are currently active. In a steady state, it also equals the growth rate of the active number of firms, which equals the growth rate of aggregate employment.

have a history of rapid enough growth to match the age distribution, even though Gibrat's law holds. The results presented in this paper show that this conjecture is wrong when shocks tend to average out within a firm. The strings of positive growth needed are too unlikely, and currently active large firms should be about 750 years old if Gibrat's law holds.

1.1. *Related literature*

This paper goes back to, interprets, and builds on the type of growth process initially proposed by Yule (1925) and Simon (1955). Yule (1925) was concerned with the number of species in biological genera, and Simon (1955) with word frequencies, city sizes and income distributions. In the context of cities, (Krugman, 1996, p. 96) described the time it takes for cities to grow large in Simon's model as an unresolved problem. Simon and Bonini (1958), Ijiri and Simon (1964) and many others since studied firm growth. Klette and Kortum (2004) describe an economy based on the quality-ladder model of Grossman and Helpman (1991) in which firm size follows a birth–death process, as in this paper. In their economy, incumbent firms cannot grow on average because there is a fixed set of commodities and new entrants continuously capture the markets for some of those commodities. This makes it impossible for large firms to arise. This difficulty is resolved here by considering an economy in which the number of commodities can grow over time, as in Romer (1990) and Young (1998). Even without growth in the number of markets, a thick-tailed size distribution can arise in the Klette and Kortum (2004) economy if Gibrat's law is relaxed along the lines described in this paper.

The models in this paper are highly tractable analytically and inevitably stylized. Lentz and Mortensen (2006) use a version of the Klette and Kortum (2004) economy with additional and more flexible sources of heterogeneity. They do not address the thin-right-tail problem but estimate their model using panel data on Danish firms.⁴ The Danish firm size data do not appear to exhibit the striking Pareto shape that is found reliably in U.S. data. The small size of the Danish economy may well account for this—there are as many firms in the U.S. as there are people in Denmark. When it comes to examining the right tail of the size distribution, a model economy with a continuum of firms could simply be a better abstraction for the U.S. than for a small country like Denmark. In addition, small countries will have fewer very large firms if the replication of blueprints across national boundaries or outside language areas comes at additional costs.

Firms in this paper are organizations that operate in (monopolistically) competitive markets and grow through continuous investment in new blueprints, at a level that is proportional to the size of the firm. One can alternatively view a firm as a trading post or network in which agents trade repeatedly. Gibrat's law and the observed size distribution arise if there is population growth and agents search for firms by randomly sampling other agents and matching with the firm with which the agent sampled is already matched. A simple version of such a model is described in Luttmer (2006). Related models of network formation are presented in Jackson (2006) and Jackson and Rogers (2007), and the extensive literature cited therein. Deciding on the relative importance of these alternative interpretations poses difficult identification problems.

1.2. *Outline*

The economy and its balanced growth path are described in Section 2, together with two alternative formulations of the role of blueprints in production. The stationary size and age distributions are derived in Section 3 and formulas are given for the tail index ζ in the Gibrat and non-Gibrat cases (Propositions 3 and 4), and for the mode of the age distribution of large firms when both

4. See also Seker (2007) for related work on Chilean establishments.

Gibrat’s and Zipf’s law hold (Section 3.5). Calibrations are in Section 4. All proofs and a description of the data are in the appendix.

2. THE ECONOMY

Blueprints are costly to replicate or produce from scratch. In the baseline version of the economy, a blueprint describes the idea for a particular final good. No two blueprints are the same, whether produced by replication or from scratch. Final goods producers are monopolistic competitors. Time is continuous and indexed by $t \in [0, \infty)$.

2.1. Consumers

There is a growing population of consumers measured by $H_t = He^{\eta t}$ at time t . The dynastic preferences of the representative consumer over aggregate consumption flows C_t are determined by

$$E_0 \left[\int_0^\infty e^{-\rho t} H_t \frac{(C_t/H_t)^{1-\gamma}}{1-\gamma} dt \right].$$

The parameters η , ρ and γ are positive and $\gamma = 1$ is interpreted as logarithmic utility. Markets are complete and consumers face standard budget constraints. The resulting interest rate in consumption numeraire is related to the consumption growth rate via

$$r_t = \rho + \gamma \left[\frac{DC_t}{C_t} - \eta \right]. \tag{1}$$

Aggregate consumption is a CES composite of differentiated commodities, as in Dixit and Stiglitz (1977),

$$C_t = \left[\int C_{\omega,t}^{1-1/\sigma} dN_t(\omega) \right]^{1/(1-1/\sigma)},$$

where $N_t(\omega)$ is the measure of type- ω commodities and $\sigma > 1$ is the elasticity of substitution. In the baseline specification, all commodity types are the same and all producers choose to charge the same price p_t . Consumers therefore set $C_{\omega,t} = C_t(p_t)$, which implies that

$$C_t = C_t(p_t) N_t^{1/(1-1/\sigma)}, \tag{2}$$

where $N_t = \int dN_t(\omega)$. Cost minimization implies that commodity demands are

$$C_t(p) = \left(\frac{p}{P_t} \right)^{-\sigma} C_t, \tag{3}$$

where P_t is the price index $P_t = p_t N_t^{-1/(\sigma-1)}$. Note that the prices of differentiated commodities and the composite good are quoted in some arbitrary numeraire. All other prices will be expressed in units of the composite commodity.

2.2. Producers

Given a blueprint for a particular differentiated commodity, a producer can use l units of labour to produce $Z_t l$ units of the commodity, where $Z_t = Ze^{\theta t}$ evolves exogenously. Given a real wage w_t , the marginal cost of one unit of a commodity is thus w_t/Z_t in units of composite good, and the constant elasticity demand curves (3) imply that producers set prices at a constant

markup over marginal cost, $p_t/P_t = (w_t/Z_t)/(1 - 1/\sigma)$. Combining this with equations (2) and (3) determines the equilibrium real wage as a function of the state (Z_t, N_t) ,

$$w_t = (1 - 1/\sigma)Z_t N_t^{1/(\sigma-1)}. \quad (4)$$

The amount of labour needed to satisfy the resulting demand for a typical commodity is $l_t = C_t(p_t)/Z_t$. Using equation (3) and the prices p_t/P_t set by producers, this gives

$$l_t = \left(\frac{(1 - 1/\sigma)Z_t}{w_t} \right)^{\sigma-1} \frac{(1 - 1/\sigma)C_t}{w_t}. \quad (5)$$

The markup $1/(1 - 1/\sigma)$ of price over marginal cost implies that profits measured in units of the composite good will be $w_t l_t / (\sigma - 1)$.

2.3. New blueprints

The producer of a differentiated commodity needs a blueprint to produce. Blueprints depreciate in a one-hoss-shay fashion at an average rate λ_t . New blueprints for distinct differentiated commodities can be produced by using labour to replicate existing blueprints, or from scratch by entrepreneurs. The respective rates at which this occurs in equilibrium are denoted by μ_t and ν_t (a mnemonic for λ_t , μ_t and ν_t is “less,” “more” and “new.”) The number of blueprints therefore evolves according to

$$DN_t = (\nu_t + \mu_t - \lambda_t)N_t. \quad (6)$$

An initial condition determines N_0 .

2.3.1. Replication of existing blueprints. A new blueprint produced from an existing blueprint arrives following an exponentially distributed waiting time with mean $\mu_t = f(i_t)$, where i_t is labour employed in the replication process. An existing blueprint is lost following an exponentially distributed waiting time with mean $\lambda_t = g(j_t)$, where j_t is labour used to “maintain” the blueprint. Note that an existing blueprint generates revenues from its use in the production of a commodity and as an input in the production of new blueprints.⁵ The value q_t of a blueprint must satisfy the Bellman equation

$$r_t q_t = \max_{\substack{\mu \leq f(i) \\ \lambda \geq g(j)}} \left\{ w_t \left(\frac{l_t}{\sigma - 1} - [i + j] \right) + (\mu - \lambda)q_t + Dq_t \right\}, \quad (7)$$

together with a transversality condition. The blueprint production function f is increasing and exhibits strictly decreasing returns to scale. The blueprint depreciation function g is assumed to be strictly decreasing and convex. For simplicity, both f and g are assumed to be sufficiently smooth, with slopes that are unbounded near zero and converge to zero for large i and j . The optimal levels of investment in new blueprints and maintenance of existing blueprints are determined by

$$\mu_t = f(i_t), \quad \lambda_t = g(j_t), \quad q_t Df(i_t) = -q_t Dg(j_t) = w_t. \quad (8)$$

The technology assumptions ensure that μ_t and $-\lambda_t$ are increasing in q_t . Blueprints are replicated more quickly and maintained better when their value is high.

5. The model of how Wal-Mart has expanded since 1962 described in Holmes (2006) has this feature. The key assumption here is that K-Mart cannot simultaneously look at a Wal-Mart blueprint to produce a new blueprint of its own. As in Boldrin and Levine (1999, 2006), and unlike Luttmer (2007), spillovers are assumed to be of secondary importance in this economy.

2.3.2. New designs by entrepreneurs. New blueprints can also be designed from scratch by agents acting as entrepreneurs, without the input of an existing blueprint. At any point in time, every agent in the economy is endowed with one unit of effort that can be allocated to two tasks: supplying labour or attempting to produce a blueprint. Every agent has a skill vector (x, y) , where x is the rate at which the agent can develop new blueprints and y is the amount of labour the agent can supply per unit of time. Comparative advantage determines occupational choice. Ignoring ties, agents with skill vectors that satisfy $q_t x > w_t y$ will choose to be entrepreneurs who design blueprints and agents with skill vectors that satisfy $q_t x < w_t y$ will choose to be employees.

There is a time-invariant talent distribution T defined over the set of all possible skill vectors, as in the Roy model of Rosen (1978). This talent distribution has a finite mean. For simplicity, it is assumed to have a density so that ties play no role. The resulting per capita supplies of entrepreneurial effort and labour are

$$E(q_t/w_t) = \int_{q_t x \geq w_t y} x dT(x, y), \quad (9)$$

$$L(q_t/w_t) = \int_{q_t x < w_t y} y dT(x, y), \quad (10)$$

respectively. Clearly, the supply of entrepreneurial effort is increasing in q_t/w_t , and the supply of labour is decreasing, both ranging between 0 and the mean skill in the population. If the talent distribution is Fréchet, then the elasticity of $E(q_t/w_t)/L(q_t/w_t)$ with respect to q_t/w_t is constant (Luttmer, 2008).

2.4. Equilibrium

Given a per capita supply of entrepreneurial effort $E(q_t/w_t)$ and a stock of blueprints N_t , the rate v_t at which new blueprints are added by entrepreneurs is determined by

$$v_t N_t = H_t E(q_t/w_t). \quad (11)$$

Labour market clearing requires that

$$(l_t + i_t + j_t) N_t = H_t L(q_t/w_t). \quad (12)$$

The equilibrium is determined by equations (1)–(12), an initial condition for N_0 and a transversality condition for $q_t N_t$.

Because the product market distortion arising from monopolistic competition is the same in all markets and at all times, and because agents supply their time inelastically, it turns out that the equilibrium allocation is Pareto efficient. It is possible to characterize the equilibrium dynamics in terms of only one state and one costate variable and use a phase diagram to construct an equilibrium that converges over time to a balanced growth path.⁶

2.5. Balanced growth

A key feature of the balanced growth path is that the allocation of labour per blueprint is constant at some (i, j, l) . Population growth then implies that the measure of blueprints is given by

6. The rate at which blueprint capital is accumulated in this economy depends intricately on the shape of the production and depreciation functions f and g and the shape of the talent distribution. Adjustment to the balanced growth path may be slow and asymmetric. A detailed analysis is beyond the scope of this paper.

$N_t = Ne^{\eta t}$ for some N . Because of equations (4) and (5), wages and per capita consumption grow at the rate $\kappa = \theta + \eta/(\sigma - 1)$. The term $\eta/(\sigma - 1)$ measures gains from variety, as in Young (1998). By equation (1), the implied interest rate is $r = \rho + \gamma\kappa$. The flow profit from producing a commodity is $w_l l/(\sigma - 1)$. It follows that $[q_t, w_t] = [q, w]e^{\kappa t}$. The Bellman equation (7) then implies that wages and blueprint prices must satisfy the present-value condition

$$\frac{q}{w} = \frac{\frac{l}{\sigma-1} - (i+j)}{r - \kappa - (\mu - \lambda)}, \quad (13)$$

where (i, j) and (μ, λ) satisfy

$$\mu = f(i), \quad \lambda = g(j), \quad (q/w)Df(i) = -(q/w)Dg(j) = 1. \quad (14)$$

Holding fixed l , these conditions imply that q/w is equal to the maximum subject to $[\mu, \lambda] = [f(i), g(j)]$ of the right-hand side of equation (13), as long as this is finite.⁷ The fact that the aggregate number of blueprints grows at the rate η implies that new blueprints must be added by entrepreneurs at the rate $\nu = \eta - (\mu - \lambda)$. If $E(q/w)$ is positive, then the entrepreneurial supply of blueprints (11) determines the steady-state supply of blueprints via

$$\frac{N}{H} = \frac{E(q/w)}{\eta - (\mu - \lambda)}. \quad (15)$$

Alternatively, $E(q/w) = 0$ and $\eta = \mu - \lambda$. Along a balanced growth path, the labor-market clearing condition (12) implies a derived demand for blueprints equal to

$$\frac{N}{H} = \frac{L(q/w)}{i + j + l}. \quad (16)$$

The balanced growth conditions (13)–(16) determine (i, j, l) , (μ, λ) , q/w and N/H . The level of wages follows from (4) and aggregate consumption can be obtained from equation (5), wages, and l .

Given a positive q/w that is not too large, the conditions (13)–(14) can be solved for the labour allocation (i, j, l) and the resulting blueprint creation and destruction rates μ and λ . It is not difficult to verify that (i, j, l) and $\mu - \lambda$ are increasing in q/w . Since $E(q/w)$ is increasing in q/w , this implies a steady-state supply of blueprints (15) that is increasing in q/w . Since $L(q/w)$ is decreasing in q/w , the derived demand for blueprints (16) is decreasing in q/w . There can therefore be at most one price q/w that clears the market for blueprints in steady state.

The replication technology must be assumed to satisfy $f(0) - g(0) < \eta$ or else $\eta \geq \mu - \lambda$ cannot hold. The assumption $r - \kappa > \eta$ ensures that $\eta \geq \mu - \lambda$ implies $r - \kappa > \mu - \lambda$. The fact that $E(q/w)$ and $L(q/w)$ go to zero as q/w goes to, respectively, zero and infinity, can now be used to argue that equations (13)–(16) does in fact have a solution. It remains to show that the decision problem of blueprint owners is well defined. This follows because blueprint owners cannot obtain unbounded profits by replicating more quickly than $r - \kappa$. The fact that $f(i) - g(j)$ is increasing and concave implies that, at the proposed equilibrium, the flow cost of doing so would exceed the flow revenues $wl/(\sigma - 1)$ per blueprint. Together, these results establish the following proposition.

7. That is, if and only if l is low enough to ensure that $\max_{i,j} \{f(i) - g(j) : i + j \leq l/(\sigma - 1)\}$ does not exceed $r - \kappa$. The value of a blueprint is infinite for l outside this range.

Proposition 1. *Suppose that $\rho + \gamma \kappa > \kappa + \eta$ and $\eta > f(0) - g(0)$. Suppose that the talent distribution is such that $E(q/w) > 0$ for all strictly positive q/w . Then, equations (13)–(16) define the unique balanced growth path and $\eta > \mu - \lambda$.*

A balanced growth path with $E(q/w) = 0$ can arise if the talent distribution has bounded support. In such an equilibrium, new blueprints are only produced using replication from an initial stock of blueprints.

2.6. Alternative blueprint interpretations

In the setup considered so far, different blueprints specify distinct differentiated commodities that are produced subject to constant returns and are sold to all consumers. The equilibrium conditions for this economy also apply to an economy in which consumers live in many different locations and blueprints are location specific. With minor modifications, the same framework can be used as well to consider competitive final goods markets and blueprints containing the specifications for production facilities or plants that are subject to decreasing returns. The following discussion elaborates on these two interpretations. They are benchmarks. Hybrid formulations are more plausible, but also less analytically tractable.

2.6.1. Sales offices or stores. Suppose that at any point in time, consumers are evenly distributed across many locations. In each location, there are many consumers who can only buy from local stores. Preferences are as in equation (2), with N_l now denoting the measure of stores in a particular location. An entrepreneur can create a blueprint for a store in a randomly selected location. The store sells a new differentiated product. The blueprint can then be copied to operate stores selling the same differentiated product in randomly selected new locations. There is an economy-wide market for labour services, or, equivalently, output is produced where workers live and can be shipped to stores at no cost.

Because there are many locations, replicated blueprints are always assigned to new locations, and every new store sells a commodity that is new to the market in which it is introduced. Assuming that there is a very large number of blueprints that can be copied, every location receives a constant flow of new stores, and stores are uniformly distributed across locations. As a result, new stores face the same market conditions everywhere.⁸ With this, the analysis proceeds as before.

2.6.2. Jobs, production lines, plants. Instead of assuming that the output of every producer is unique, suppose there is one competitive market for final goods. A blueprint defines a particular job, production line or plant that is subject to decreasing returns to scale. Each plant can use l_l units of labour to produce output $Z_l F(1, l_l)$ for some constant returns to scale production function F . Growth in variety is no longer a source of consumption growth. Along a balanced growth path, wages grow at the same rate as Z_l , the amount of labour used per blueprint is constant, and the number of blueprints grows at the population growth rate η . In contrast to the Dixit–Stiglitz formulation used elsewhere in this paper, no constant elasticity assumptions are

8. There must be many more stores than locations. Imagine markets are non-overlapping intervals of length $1/A$ in $[0, 1]$, where $A \in \mathbb{N}$. Each one of the A markets has τA consumers and there are σA^2 stores that are randomly assigned to points in $[0, 1]$. The ratio of stores to consumers is σ/τ . As A becomes large, the proportion of all stores assigned to the region $[0, x]$ converges to x . If the number of stores were σA instead, then the number of stores in different markets would remain random and converge to a Poisson distribution. Market conditions would vary across locations, and strategic considerations would come into play in each market.

needed. The production function F is general, even though Z_t is not labour augmenting in the usual sense. In units of blueprints, the per capita capital stock is constant. But the market value of the capital stock and the cost of producing new capital grows at the same rate as Z_t , wages, and output per capita.

3. THE DISTRIBUTION OF FIRM SIZE AND AGE

The economy described up to now has agents who consume, supply labour, and act as entrepreneurs. Everyone can own blueprints and there are no firms. A transaction cost argument can be used to motivate a definition of what firms are in this economy.

Consider an entrepreneur who has just developed a new blueprint. To hire labour to produce the associated commodity and develop further copies of the same blueprint, the entrepreneur can set up a firm at no cost. This defines a firm entry. Claims to firms can be traded freely. But there is a cost, potentially very small, involved in firms hiring entrepreneurs to develop new blueprints from scratch, in selling blueprints to firms, and in merging firms. There are no cost advantages to any of these transactions, and so they will not occur in equilibrium.⁹ A firm will therefore only gain new blueprints through “organic growth” by replicating its existing blueprints. A firm only loses blueprints as they depreciate at the rate λ .¹⁰ Exit occurs when a firm has lost all its blueprints.

The measure of firms with n blueprints at time t is denoted by $M_{n,t}$. The aggregate measure of blueprints is therefore

$$N_t = \sum_{n=1}^{\infty} n M_{n,t}. \quad (17)$$

Over time, the change in the number of firms with one blueprint is

$$DM_{1,t} = \lambda 2M_{2,t} + \nu N_t - (\mu + \lambda)M_{1,t}, \quad (18)$$

where μ , λ and $\nu = \eta - (\mu - \lambda)$ are equilibrium rates that are constant along the balanced growth path. The number of firms with one blueprint increases because firms with two blueprints lose one or because of entry. The number declines because firms with one blueprint gain or lose a blueprint. Similarly, the numbers of firms with more than one blueprint evolve according to

$$DM_{n,t} = \mu(n-1)M_{n-1,t} + \lambda(n+1)M_{n+1,t} - (\mu + \lambda)nM_{n,t}, \quad (19)$$

for all $n - 1 \in \mathbb{N}$. The joint dynamics of N_t and $\{M_{n,t}\}_{n=1}^{\infty}$ is fully described by equations (17)–(19).

3.1. The stationary size distribution

Along the balanced growth path, N_t grows at the rate η and a stationary firm size distribution exists if equations (17)–(19) have a solution that satisfies $DM_{n,t} = \eta M_{n,t}$ for all $n \in \mathbb{N}$. Given that N_t and $M_{n,t}$ grow at the common rate η , one can then define

$$P_n = \frac{M_{n,t}}{\sum_{k=1}^{\infty} M_{k,t}}$$

9. Of course, these transactions do occur in the data. This is a familiar and important failure of the type of model described in this paper. Chatterjee and Rossi-Hansberg (2006) provide an interesting model of firm size in which adverse selection makes it difficult for firms to hire entrepreneurs.

10. Bernard, Redding and Schott (2006) document the importance of turnover in the mix of products sold by U.S. manufacturing firms. They report that less than 1% of product adds and drops are associated with mergers or acquisitions.

for all $n \in \mathbb{N}$. This is the fraction of firms with n blueprints. It is analytically more convenient to use the fraction of all blueprints held by firms of size n . This is given by

$$Q_n = \frac{nM_{n,t}}{\sum_{k=1}^{\infty} kM_{k,t}}$$

for all $n \in \mathbb{N}$. With these definitions, equation (18) becomes

$$\eta Q_1 = \lambda Q_2 + \nu - (\mu + \lambda) Q_1, \tag{20}$$

and equation (19) implies

$$\frac{1}{n} \eta Q_n = \mu Q_{n-1} + \lambda Q_{n+1} - (\mu + \lambda) Q_n, \tag{21}$$

for $n - 1 \in \mathbb{N}$. Note that these equations only depend on the parameters μ/η and λ/η . The stationary distribution cannot depend on the units in which time is measured.

Proposition 2. *Suppose that η, μ, λ and $\nu = \eta - (\mu - \lambda)$ are positive. Define the sequence $\{\beta_n\}_{n=0}^{\infty}$ by the recursion $\beta_n = 1/(1 - (\mu\beta_{n-1}/\lambda) + (\eta + \mu n)/\lambda n)$ and the initial condition $\beta_0 = 0$. This sequence is monotone and converges to $\min\{1, \lambda/\mu\}$. The only non-negative and summable solution to equations (20)–(21) is given by*

$$Q_n = \frac{\nu}{\mu} \sum_{k=0}^{\infty} \frac{1}{\beta_{n+k}} \left(\prod_{m=n}^{n+k} \beta_m \right) \prod_{m=1}^{n+k} \frac{\mu\beta_m}{\lambda}. \tag{22}$$

If $\mu \neq \lambda$, then

$$Q_n \sim \frac{\nu}{|\mu - \lambda|} \prod_{m=1}^{n-1} \frac{\mu\beta_m}{\lambda}. \tag{23}$$

Here, equation (23) means that the ratio of the left- and right-hand sides converges to 1 as n becomes large. If $\nu = 0$, then the only non-negative and summable solution to equations (20)–(21) is identically zero, implying that there does not exist a stationary in this case. If $\nu > 0$, then equation (22) adds up to 1 by construction and defines a stationary size distribution $\{P_n\}_{n=1}^{\infty}$ via $P_n \propto Q_n/n$. The mean firm size can be written as $1/(\sum_{n=1}^{\infty} Q_n/n)$, and this is also finite by construction. Appendix B proves these results and gives an explicit solution for Q_n in the more general case that arises when the size distribution of entrants is non-degenerate.

When $\lambda > \mu$, the properties of the right-hand side of equation (23) are very different from what they are when $\mu > \lambda$. If $\lambda > \mu$, then Q_n is bounded above by a multiple of the geometrically declining sequence $(\mu/\lambda)^n$. On the other hand, if $\mu > \lambda$, then $\mu\beta_n/\lambda \uparrow 1$, and hence the right-hand side of equation (23) declines at a rate that is slower than any given geometric rate. The following proposition gives a further characterization of the right tail of the distribution.

Proposition 3. *Suppose that $\eta > \mu - \lambda > 0$. Then, the right tail probabilities $R_n = \sum_{k=n}^{\infty} P_k$ of the stationary firm size distribution satisfy*

$$\lim_{n \rightarrow \infty} n \left(1 - \frac{R_{n+1}}{R_n} \right) = \zeta,$$

where $\zeta = \eta/(\mu - \lambda)$. That is, R_n is a regularly varying sequence with index $-\zeta$.

This means that $\lim_{n \rightarrow \infty} R_{[xn]}/R_n = x^{-\zeta}$ for any $x > 0$, where $[xn]$ is the integer part of xn . An implication is that $n^z R_n \rightarrow 0$ for all $z < \zeta$ and $n^z R_n \rightarrow \infty$ for all $z > \zeta$. Even though this does not describe precisely what happens to $n^\zeta R_n$ for large n , the parameter ζ will continue to be referred to as the tail index of the size distribution.¹¹ The limiting tail index $\zeta = 1$ associated with Zipf's law arises when the rate $\nu = \eta - (\mu - \lambda)$ at which blueprints are introduced by entrepreneurs converges to zero. Appendix B shows that Proposition 3 holds more generally if the size distribution of entrants has a right tail that is regularly varying with an index smaller than $-\zeta$.

For comparison, consider the economy of Klette and Kortum (2004). There, $\eta = 0$ and $\mu < \lambda$. This turns equations (20)–(21) into a linear difference equation with constant coefficients that is easy to solve. The resulting firm size distribution is R.A. Fisher's logarithmic series distribution, which has $P_n \propto (\mu/\lambda)^n/n$. As a result, right tail probabilities converge to zero even more quickly than a geometric sequence. To generate a thick right tail, firms must grow on average, and in the economy described here this requires population growth. A tail index ζ close to 1 can only arise if growth in the number of blueprints is mostly due to incumbents rather than new entrants. It is critical that firms grow exponentially. If firms accumulate new blueprints at some constant rate μ , instead of μn , then the size distribution would be Poisson like, with a geometrically bounded right tail.

3.2. Firm entry and exit rates

The flow of blueprints introduced by new firms is νN_t . Each new firm starts with one blueprint, and so νN_t is also the flow of new firms that enter per unit of time. The firm entry rate is therefore $\varepsilon = \nu N_t/M_t$, where $M_t = \sum_{n=1}^{\infty} M_{n,t}$ is the number of firms in the economy. It follows that $\varepsilon/\nu = N_t/M_t = \sum_{n=1}^{\infty} n P_n$ is the average firm size. An alternative way to calculate the firm entry rate ε is to note that the only firms that exit in this economy are firms with one remaining blueprint. The proportion of such firms is P_1 , and they exit at a rate λ . The resulting balance $\varepsilon - \lambda P_1$ of firms entering and exiting per unit of time must equal the rate η at which the number of firms grows over time. These two calculations can be summarized as

$$\varepsilon = \nu \sum_{n=1}^{\infty} n P_n = \eta + \lambda P_1. \quad (24)$$

Just like the blueprint entry rate $\nu + \mu = \eta + \lambda \geq \eta$, the firm entry rate can be no less than the population growth rate, and this lower bound is attained only when firms never lose blueprints and therefore never exit. The two equations given in equation (24) and $Q_1 = P_1/\sum_{n=1}^{\infty} n P_n$ imply $\varepsilon/\eta = \nu/(\nu - \lambda Q_1)$. Together with equation (22), this yields an explicit formula for the firm entry rate relative to the population growth rate. In turn, this implies an explicit formula for the average firm size ε/ν .

3.3. Firm type transitions

The data shown in Figure 1 suggest that some firms initially grow at rates that far exceed the bound $\mu - \lambda < \eta$ implied by Proposition 1 and that these growth rates decline with firm size and age. A simple way to account for this slowdown and examine its implications for the stationary size distribution is as follows. Suppose there are high- and low-quality blueprints. High-quality blueprints imply a productivity $Z_H e^{\theta t}$ and low-quality blueprints imply a productivity $Z_L e^{\theta t}$,

11. See Bojanic and Seneta (1973) for the definition of regularly varying sequences and some of its implications. Bingham, Goldie and Teugels (1987) is a useful source on the general topic of regular variation.

where $Z_H > Z_L$. Entrepreneurs produce high-quality blueprints with probability $\alpha \in (0, 1]$ and low-quality blueprints with probability $1 - \alpha$.¹² Incumbent firms replicate blueprints as before, preserving their quality. But high-quality firms transition to become low-quality firms following independent and exponentially distributed waiting times with a mean $1/\delta_H$. When such a transition happens, all blueprints of the firm turn into low-quality blueprints, permanently. Any new blueprints created by the firm thereafter will be of low quality. Low-quality firms can also exit randomly at a rate δ_L , irrespective of the number of blueprints that make up the firm.

One possible interpretation for these kinds of firm type transitions is that some aspect of the environment for which the initial blueprint of a firm was created changes permanently. Outside the formal model described here, a relative decline in the quality of a firm's blueprints could arise from competing firms catching up. An alternative interpretation for the decline in firm growth rates is that blueprints are location specific and that firms initially implement blueprints in the most profitable locations.

Along a balanced growth path, the present-value condition (13) must be modified to account for the loss in value that occurs when a blueprint transitions from one type to another. The incentives to replicate and maintain continue to be determined by equation (14). Let $\bar{q} = \alpha q_H + (1 - \alpha)q_L$, where q_H and q_L are the respective prices of high- and low-quality blueprints. The steady-state number of high-quality blueprints is $N_H/H = \alpha E(\bar{q}/w)/(\eta + \delta_H - [\mu_H - \lambda_H])$. Low-quality blueprints are produced by entrepreneurs, by incumbent replication, and because a flow of high-quality blueprints depreciate in quality. This implies $N_L/H = [(1 - \alpha)E(\bar{q}/w) + \delta_H N_H/H]/(\eta + \delta_L - [\mu_L - \lambda_L])$. Write (i_1, j_1, l_1) for the allocation of labour to a type-I blueprint. As long as the talent distribution for entrepreneurs is unbounded, $E(\bar{q}/w)$ is positive and the labour market clears if

$$\frac{L(\bar{q}/w)}{E(\bar{q}/w)} = \frac{(1 - \alpha)(i_L + j_L + l_L)}{\eta + \delta_L - (\mu_L - \lambda_L)} + \frac{\alpha}{\eta + \delta_H - (\mu_H - \lambda_H)} \left(i_H + j_H + l_H + \frac{\delta_H(i_L + j_L + l_L)}{\eta + \delta_L - (\mu_L - \lambda_L)} \right).$$

As in the case of equations (15)–(16), labour market clearing forces $\eta + \delta_H > \mu_H - \lambda_H$ and $\eta + \delta_L > \mu_L - \lambda_L$ in any equilibrium in which entrepreneurs contribute to the supply of new blueprints.

One can verify that $Z_H > Z_L$ implies $i_H > i_L$, $j_H > j_L$, $l_H > l_L$ and $q_H > q_L$. While their quality advantage lasts, high-quality firms have stronger incentives to replicate and maintain blueprints than low-quality firms. High-quality firms choose to grow faster than low-quality firms. This can account for the thick tail of the size distribution.

Proposition 4. *Suppose some firms enter as high-quality firms and transition to low-quality firms at a positive rate δ_H . Low-quality firms also exit randomly at a rate δ_L . Then, along the balanced growth path, $\mu_H - \lambda_H > \mu_L - \lambda_L$, $\eta + \delta_H > \mu_H - \lambda_H$ and $\eta + \delta_L > \mu_L - \lambda_L$. The stationary size distribution has a tail index ζ given by*

$$\zeta = \min \left\{ \frac{\eta + \delta_H}{[\mu_H - \lambda_H]^+}, \frac{\eta + \delta_L}{[\mu_L - \lambda_L]^+} \right\}.$$

The right tail of the size distribution declines geometrically if this is infinite.

12. Alternatively, one can assume that entrepreneurs have potentially different skills for producing high- and low-quality blueprints. Relative prices and comparative advantage then determine the quality mix of start-up blueprints.

The actual size distribution and a proof of Proposition 4 are implied by the results in Appendix B. If $\zeta = (\eta + \delta_H)/(\mu_H - \lambda_H)$, then large firms arise because of the rapid growth of new firms. This can generate a thick tail even if there is no population growth.

3.4. Firm age and size

The age distribution among large firms is a useful tool for assessing alternative interpretations of the firm size distribution. This age distribution can be constructed from the size distributions of cohorts of firms that are the same age.

3.4.1. The size distribution of a cohort. Consider a cohort of firms initially with $k \in \mathbb{N}$ blueprints and in a common growth phase. While in this initial growth phase, firms of size $n \in \mathbb{N}$ gain blueprints at the rate μn and lose blueprints at the rate λn . Firms in this cohort also transition randomly into a new growth phase at a rate δ . Let $\mathcal{T}_{-1,k}(a)$ denote the fraction of firms in the cohort that have made this transition by age a . Define $\mathcal{T}_{n,k}(a)$ to be the fraction of firms that have $n \in \mathbb{N}$ blueprints at age a and have not made the transition. Let $\mathcal{T}_{0,k}(a)$ represent the firms in the cohort that have exited as a result of losing their last blueprint. Only firms that have not yet exited can transition into a new growth phase,

$$D\mathcal{T}_{-1,k}(a) = \delta[1 - \mathcal{T}_{-1,k}(a) - \mathcal{T}_{0,k}(a)]. \quad (25)$$

Exit occurs when a firm loses its last blueprint, and hence

$$D\mathcal{T}_{0,k}(a) = \lambda\mathcal{T}_{1,k}(a). \quad (26)$$

The number of firms with n blueprints and still in the original growth phase by age a must satisfy

$$D\mathcal{T}_{n,k}(a) = \mu(n-1)\mathcal{T}_{n-1,k}(a) + \lambda(n+1)\mathcal{T}_{n+1,k}(a) - [\delta + (\mu + \lambda)n]\mathcal{T}_{n,k}(a) \quad (27)$$

for all $n \in \mathbb{N}$. Note that the $-\delta\mathcal{T}_{n,k}(a)$ term is not scaled by n , reflecting the assumption that the transition to a new growth phase is independent of size.

Proposition 5. For any $\mu > 0$ and $\lambda \geq 0$ define $\gamma(a) = (e^{(\mu-\lambda)a} - 1)/(e^{(\mu-\lambda)a} - \lambda/\mu)$. Fix some $k \in \mathbb{N}$. Then, equations (25)–(27) together with the initial condition $\mathcal{T}_{k,k}(0) = 1$ give

$$\mathcal{T}_{-1,k}(a) = \delta \int_0^a e^{-\delta b} \left(1 - \left[\frac{\lambda}{\mu} \gamma(b) \right]^k \right) db$$

and

$$\mathcal{T}_{0,k}(a) = \lambda k \int_0^a e^{-\delta b} \left[1 - \frac{\lambda}{\mu} \gamma(b) \right] \left[\frac{\lambda}{\mu} \gamma(b) \right]^{k-1} [1 - \gamma(b)] db,$$

as well as

$$\mathcal{T}_{n,k}(a) = e^{-\delta a} \sum_{m=1}^{\min\{n,k\}} \binom{k}{m} \binom{n-1}{m-1} \left[1 - \frac{\lambda}{\mu} \gamma(a) \right]^m \left[\frac{\lambda}{\mu} \gamma(a) \right]^{k-m} [1 - \gamma(a)]^m \gamma^{n-m}(a),$$

for all $n \in \mathbb{N}$.

For $\delta = 0$ and $k = 1$, this solution can be found in Klette and Kortum (2004). The probability generating function for $\delta = 0$ and $k \in \mathbb{N}$ is in Kendall (1948). Using the fact that $\gamma(a)$ goes to zero

as age goes to zero one can verify that $\mathcal{T}_{k,k}(a) \uparrow 1$ as age goes to zero. The solution for $\mathcal{T}_{0,k}(a)$ follows directly from $\mathcal{T}_{1,k}(a)$ and integrating equation (26). Summing $\mathcal{T}_{n,k}(a)$ over all $n \in \mathbb{N}$ gives $1 - \mathcal{T}_{-1,k}(0) - \mathcal{T}_{0,k}(0) = e^{-\delta a} (1 - [\gamma(b)\lambda/\mu]^k)$ and then $\mathcal{T}_{-1,k}(0)$ follows from integrating equation (25). The proof of Proposition 5 can be completed by computing the derivative of $\mathcal{T}_{n,k}(a)$ and checking equation (27) for any $n \in \mathbb{N}$. Appendix C gives a more constructive proof based on the observation that, conditional on no transition, a firm with n blueprints gains and loses blueprints with the same probabilities as does the aggregate of n independent firms with one blueprint each.

If $\delta = 0$, then $\mathcal{T}_{0,k}(a) \rightarrow \min\{1, \lambda/\mu\}$ as the age of a cohort grows without bound. If $\mu < \lambda$, then virtually all of a cohort of firms will have exited the economy after a sufficiently long time. On the other hand, if $\mu > \lambda$, then a fraction $1 - \lambda/\mu$ of any cohort of firms survives and grows forever, giving rise to a thick-tailed size distribution.

3.4.2. Age given size. Now consider the setup of Proposition 4, with a fraction α of a cohort of new firms entering with high-quality blueprints. Write $\mathcal{T}_{H,n,k}(a)$ and $\mathcal{T}_{L,n,k}(a)$ for the solutions to equations (25)–(27) associated with the parameters $(\mu_H, \lambda_H, \delta_H)$ and $(\mu_L, \lambda_L, \delta_L)$, respectively. Let $n = -1$ now represent the absorbing state low-quality firms enter into at a rate δ_L . Then, the cohort size distribution $\{p_n(a)\}_{n=-1}^\infty$ at age a is given by

$$p_n(a) = (1 - \alpha)\mathcal{T}_{L,n,1}(a) + \alpha \left(\mathcal{T}_{H,n,1}(a) + \delta_H \int_0^a \left[\sum_{k=1}^\infty \mathcal{T}_{L,n,k}(b)\mathcal{T}_{H,k,1}(a-b) \right] db \right) \quad (28)$$

for all $n + 1 \in \mathbb{N}$. For $n = -1$, the term $\alpha\mathcal{T}_{H,n,1}(a)$ drops out since high-quality firms do not transition directly into the state $n = -1$. The infinite sum on the right-hand side of equation (28) can be calculated explicitly, as reported in Appendix C. The first two terms on the right-hand side of equation (28) account for the firms that are still in their original growth phase. A flow $\delta_H\mathcal{T}_{H,k,1}(a-b)$ of high-quality firms with k blueprints transition to become low-quality firms at age $a-b$. Adding up over all sizes and ages and accounting for subsequent firm growth gives the third term. Note well that only a fraction $1 - p_{-1}(a) - p_0(a)$ of the cohort survives until age a as active firms.

Along a balanced growth path, the measure of entering firms is growing at a rate η . Consider the population of all firms that have ever entered up to a particular point in time, including those that have since exited. The exponential rate η at which the size of entering cohorts grows implies that this population has an exponential age distribution with density $\eta e^{-\eta a}$. Because $\{p_n(a)\}_{n=-1}^\infty$ includes firms that have exited, the joint density of age and size is $\eta e^{-\eta a} p_n(a)$ among all firms that have ever entered. The age density among all firms of size at least n is therefore

$$h_n(a) = \frac{e^{-\eta a} \sum_{k=n}^\infty p_k(a)}{\int_0^\infty e^{-\eta b} \sum_{k=n}^\infty p_k(b) db}. \quad (29)$$

In particular, for $n = 1$, this defines the age density among all surviving firms.

3.5. *Gibrat and Zipf: A convenient limiting case*

Consider again the economy in which all firms have the same growth parameters μ and λ . As noted after Proposition 1, if the entrepreneurial talent distribution has bounded support, then the economy may have a balanced growth path with $\eta = \mu - \lambda$. Along such a balanced growth path, there is no entry by new firms, incumbent firms grow randomly, and there is no stationary size distribution.

Alternatively, suppose the entrepreneurial talent distribution does have unbounded support, implying $\eta > \mu - \lambda$ and the existence of a stationary size distribution along the balanced growth path. Consider a sequence of such economies for which $\mu - \lambda$ approaches η from below. Suppose that η is positive and bounded away from zero, so that the tail index $\zeta = \eta/(\mu - \lambda)$ converges to 1. Suppose further that λ is also positive and bounded away from zero along the sequence. Then, the recursion (21) for $Q_n \propto nP_n$ converges to a limiting recursion that can be written as $P_n = \frac{\lambda}{\mu}(P_{n+1} + X_{n+1})$, where $X_{n+1} = \left(\frac{n-1}{n+1}\right)X_n$ for all $n-1 \in \mathbb{N}$. Note that $X_{n+1} = 2X_2/[n(n+1)]$ for all $n \in \mathbb{N}$. Iterating forward on the recursion for P_n and requiring the resulting P_n to add up to 1 then yields

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

Thus, the sequence of stationary size distributions has a well-defined limiting distribution. The right-tail probabilities of this limiting distribution satisfy

$$\lim_{n \rightarrow \infty} n \sum_{k=n}^{\infty} P_k = \lim_{n \rightarrow \infty} \frac{1}{\ln(\mu/\eta)} \sum_{m=0}^{\infty} \frac{n}{n+m} \left(\frac{\lambda}{\mu}\right)^{m+1} = \frac{\mu/\eta - 1}{\ln(\mu/\eta)}, \quad (30)$$

by the dominated convergence theorem. Thus, the right-tail probabilities behave like $1/n$ —they satisfy Zipf's law.

This limiting size distribution implies a limiting firm entry rate $\varepsilon = \eta + \lambda P_1 > \eta > 0$, even though the rate $\nu = \eta - (\mu - \lambda)$ at which new blueprints are introduced by new firms converges to zero. The average firm size ε/ν grows without bound, as expected from equation (24). Evaluating P_1 and using $\lambda = \mu - \eta$ gives

$$\frac{\varepsilon}{\eta} = \frac{\mu/\eta - 1}{\ln(\mu/\eta)}. \quad (31)$$

When the average incumbent firm is very large, virtually all blueprints are created by incumbent firms, even though the firm entry rate is strictly positive. Under such circumstances, equation (31) allows one to infer μ and λ simply from the population growth rate η and the firm entry rate ε .

The formulas (30) and (31) provide a simple way to infer the number of employees per blueprint. A comparison of equations (30) and (31) shows that $\varepsilon/\eta \approx nR_n$ for large n . The higher the entry rate, the more firms there will be in the right tail. If R_n is measured by the fraction of all firms with more than a certain large number of employees, then $(\varepsilon/\eta)/R_n$ provides an estimate of the associated number of blueprints n , and this then implies an estimate for the number of employees assigned to each blueprint.

Given only one growth phase, the expression (29) for the density of firm age given a size of at least n blueprints reduces to $h_n(a) \propto e^{-\eta a} \sum_{k=n}^{\infty} \mathcal{T}_{k,1}(a) = e^{-\eta a} [1 - (\lambda/\mu)\gamma(a)] \gamma^{n-1}(a)$, where $\gamma(a)$ is given in Proposition 5. In the Zipf limit $\eta = \mu - \lambda$, this depends on a only via $e^{\eta a}$. Setting the derivative of $h_n(a)$ equal to zero then shows that the mode a_{mode} of this density satisfies

$$\eta a_{\text{mode}} = \ln \left(1 + \frac{n-1}{\mu/\eta} \right) \quad (32)$$

for all $n \in \mathbb{N}$. Letting $\lambda \downarrow 0$ so that $\mu \downarrow \eta$ gives the Zipf asymptote of the Yule process, and this yields $a_{\text{mode}} = \ln(n)/\eta$. In this limit, the mode of the age density among firms of size no less than n turns out to be equal to the time it would take a firm to grow from 1 to n at a deterministic rate $\mu = \eta$. Increasing μ subject to $\eta = \mu - \lambda$ increases the variability of firm growth rates and spreads out the size distribution of an age cohort. This makes it more likely that surviving firms of a given age are large, which lowers the modal age of large firms. In fact, formula (32) shows

that the modal age of large firms converges to zero as μ/η increases without bound. But equation (31) implies that then the firm entry rate also grows without bound. This will be the tension that makes it hard to hold on to Gibrat's law.

4. U.S. EMPLOYER FIRMS

U.S. Internal Revenue Service statistics contain more than 26 million corporations, partnerships and non-farm proprietorships. Business statistics collected by the U.S. Census consist of both non-employer firms and employer firms. In 2002, there were more than 17 million non-employer firms, many with very small receipts, and close to 6 million employer firms.

Here, Census data on employer firms assembled by the U.S. Small Business Administration (SBA) will be considered. For employer firms, part-time employees are included in employee counts, as are executives. But proprietors and partners of unincorporated businesses are not (Armington, 1998, p. 9). This is likely to create significant distortions in measured employment for small firms. The SBA reports firm counts for 24 size categories, ranging from 1 to 4 employees to 10000 and more employees, as well as the number of employer firms that have no employment in March but some employment at other times during the year. Over the period 1989–2006, SBA data show that the number of firms grows roughly at the population growth rate of about 1% per annum, in line with the theory presented here and in Luttmer (2007).

Newly collected data on the age of firms with more than 10000 employees in 2008 will also be used. Two measures of firm age are reported in Figure 3 below. One is based on the date a firm was incorporated. Corporate restructuring can cause this measure of age to be much below the age of the underlying organization that constitutes the firm. An alternative measure uses the earliest date a firm or any of its components are known to have been in operation. A more detailed description of how this data was collected is given in Appendix A. Clearly, the complicated genealogy of many large corporations is not captured by the models described in this paper.

4.1. *Gibrat implies 750-year-old firms*

Panels (i) and (ii) of Figure 2 show the fitted employment size distribution assuming that there is only one growth phase. The fractions $\#\{\text{firms with employment} \leq x\}/\#\{\text{all firms}\}$ and $\#\{\text{firms with employment} \geq x\}/\#\{\text{all firms}\}$ observed in the data are displayed after merging the category of employer firms with no employment in March with the category of 1 to 4 employees. The right tail of the size distribution, shown in panel (ii), is clearly well approximated by $x^{-\zeta}$, and the slope of the log tail probabilities with respect to x is about $\zeta \approx 1.05$. Note that this estimate does not depend on the units in which firm size is measured. The formula for the tail index ζ combined with a 1% population growth rate implies that incumbent firms grow at a rate $\mu - \lambda = \eta/\zeta \approx 0.95\%$ per annum.

To decompose $\mu - \lambda$, consider first the Yule process obtained by setting $\lambda = 0$ and $\mu = 0.0095$. The only remaining free parameter is then the number of employees per blueprint $i + j + l$. To see how this parameter can be identified, write $\ell = i + j + l$ and recall that R_n is the fraction of firms with n or more blueprints. The fraction of firms with at least $x \in \{\ell n : n \in \mathbb{N}\}$ employees is then $R_{x/\ell}$. The vertical axis of panel (ii) of Figure 2 shows $\ln(R_{x/\ell})$ and the horizontal axis shows $\ln(x) = \ln(\ell) + \ln(x/\ell)$. Thus, an increase in employment per blueprint moves the model prediction $[\ln(x), \ln(R_{x/\ell})]$ to the right by the change in $\ln(\ell)$, for every $\ln(R_{x/\ell})$. Since the empirical counterparts to $[\ln(x), \ln(R_{x/\ell})]$ are pretty much on a straight line for all firms with more than 10 employees, the choice of ℓ will simultaneously either fit or fail to fit all right tail probabilities for firms with more than 10 employees. The close fit of the right tail shown in Panel

(ii) is obtained by setting $i + j + l = 2$. Panel (i) shows that the left tail is also well approximated. The stationary size distribution of a Yule process fits the empirical size distribution quite well. But a Yule process predicts no exit and a firm entry rate equal to $\varepsilon = \eta$ or about 1% per annum. Instead, the SBA reports a firm entry rate of about 11% per annum over the period 1989–2006. Actual firms do decline and exit, and entry rates are much higher than the population growth rate.

To match the evidence on firm entry, consider first the $\zeta \downarrow 1$ approximation. Given $\eta = 0.01$ and $\varepsilon = 0.11$, solving equation (31) for μ and $\lambda = \mu - \eta$ gives $\mu = 0.4216$ and $\lambda = 0.4116$. In 2008, close to 1000 firms out of a total of 6 million employer firms had at least 10 000 employees. The combination of equations (30) and (31) therefore gives $i + j + l \approx 10\,000 \times R_n / (\varepsilon / \eta) \approx 10 \times (1/6) / (0.11/0.01) = 0.15$. The more precise estimates obtained for $\zeta = 1.05$ are very similar. Increasing μ and λ subject to the constraint $\mu - \lambda = \eta / \zeta \approx 0.0095$ until the implied entry rate ε reaches the 0.11 value observed in the data yields $\mu = 0.4095$ and $\lambda = 0.4000$. Choosing the number of employees per blueprint to match the right tail probabilities now implies $i + j + l = 0.20$. The associated left and right tails are shown in Panels (i) and (ii) of Figure 2. The increased transition probabilities μ and λ raise the variance $(\mu + \lambda)/n$ of the growth rate of a firm with n blueprints, and this implies that surviving firms are more likely to have many blueprints. Fitting the right tail of the employment distribution therefore requires fewer employees per blueprint than in the case of a Yule process. But then the left tail of the size distribution no longer fits well. The higher variance cuts down, too much, on the number of small firms—they either exit or grow large.

The age distributions displayed in the upper panel of Figure 3 show a much more dramatic failure of the one-phase model of firm growth. At $\mu = 0.4095$, $\lambda = 0.4000$ and $i + j + l = 0.20$, the median age of firms with more than 10 000 employees is about 750 years. The Yule process fitted above implies a median large firm that is a couple of centuries older still. In the data, the median age of these large firms is only about 75 years. Given Gibrat's law, all firms grow at the same average rate $\mu - \lambda$, and this cannot exceed η , or about 1% per year. Deterministic growth would imply that it takes $\ln(50\,000)/0.01 \approx 1082$ years to reach the size of 10 000 employees. The $\zeta \downarrow 1$ approximation (31) for the entry rate gives $\mu/\eta \approx 42.16$ and then the approximation (32) for the mode of the age distribution of large firms gives $a_{\text{mode}} = 100 \times \ln(1 + 49\,999/42.16) \approx 708$ years. Adding variability lowers the age of the typical large firm, but the amount of variability that can be added is constrained by the entry and exit evidence.

SBA data for the period 1989–2006 show that the annual exit rate of firms with 500 or more employees is about 2.5%. As calibrated so far, the model implies that this number should be essentially zero. The annual growth rate of a firm with 500 employees has a standard deviation of only $\sqrt{(0.4095 + 0.4000)/2500} \approx 0.018$ and firms only exit after losing all blueprints. To account for the observed exit of large firms, suppose a firm may not only exit after losing its last blueprint but also randomly at a rate $\delta = 0.02$. The approximation $\zeta \downarrow 1$ then implies that surviving firms grow at a rate $\mu - \lambda = \eta + \delta$ of about 3% per year. The entry approximation (31) obtained by replacing η with $\eta + \delta$ yields $\mu/(\eta + \delta) \approx 9.1$, and hence $\mu \approx 0.2730$ and $\lambda \approx 0.2430$. With a tail index $\zeta = 1.05$ instead of $\zeta = 1$, this becomes $\mu = 0.2700$ and $\lambda = 0.2415$. Given that random exit now accounts for 2% of the 10% exit rate, randomness at the blueprint level must be lower than before. As a result, there will be more small firms, and adjusting the number of employees per blueprint to match the right tail of the size distribution now gives $i + j + l \approx 0.6$. Thus, a firm with 10 000 employees has about 16 667 blueprints, and the approximation (32) then implies $a_{\text{mode}} = \ln(1 + 16\,666/9.16)/0.03 \approx 250$, a drastic improvement over the calibration without random exit. The entire age distribution is shown in Figure 3. Deterministic growth conditional on survival would give $\ln(50\,000)/0.03 \approx 361$ or $\ln(16\,667)/0.03 \approx 324$, and so much of the reduction in age is simply due to the fact that survivors can now grow at

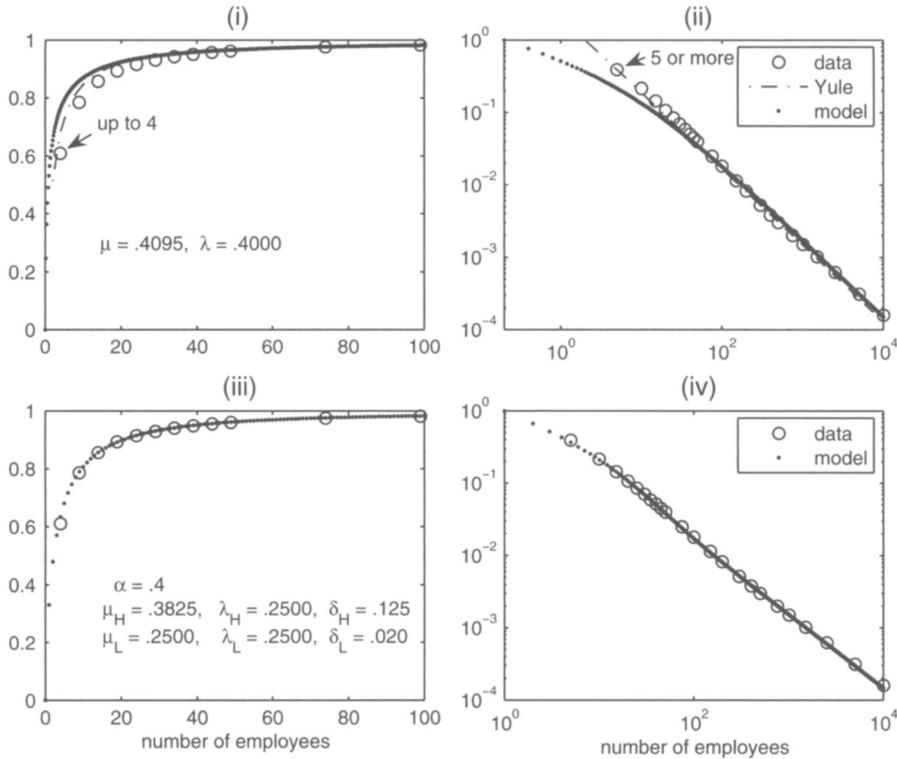


FIGURE 2
Left and right cumulative distribution functions

a 3% annual rate instead of a 1% annual rate. But this calibration still predicts that the median U.S. firm is older than the U.S. itself.

4.2. Rapid initial growth

As Figure 1 suggests, many large firms became large during relatively short periods of growth at rates far exceeding the sum of the population growth rate and the exit rate of large firms. This can account for the fact that the median large firm is only 75 years old. Proposition 4 indicates how this can also be made consistent with the observed right tail of the size distribution. Firms can grow initially at a high rate $\mu_H - \lambda_H$ and then transition at a rate δ_H to a regime with a growth rate $\mu_L - \lambda_L$ that must be below $\eta + \delta_L$. If the tail index is determined by the effects of initial rapid growth, then $\zeta = (\eta + \delta_H) / (\mu_H - \lambda_H)$. Given $\zeta \approx 1.05$ and $\eta \approx 0.01$, this implies that $\mu_H - \lambda_H$ must be close to δ_H . An initial phase with very rapid growth is possible as long as this phase is of sufficiently short average duration. The realized durations of the high-growth regime are exponentially distributed, implying that some firms grow rapidly for much longer than the average duration. This results in relatively young large firms.

Allowing for an initial growth phase adds the parameters α , μ_H , λ_H , δ_H and $i_H + j_H + l_H$. This gives more than enough flexibility to match the observed median age of large firms. The theory of Section 2 implies $i_H + j_H + l_H > i_L + j_L + l_L$ but is silent on the magnitude of the difference. Measured employment per blueprint may not reflect effective labour used per blueprint if workers differ in ability. Panels (iii) and (iv) of Figure 2 and the lower panel of Figure 3 show the size

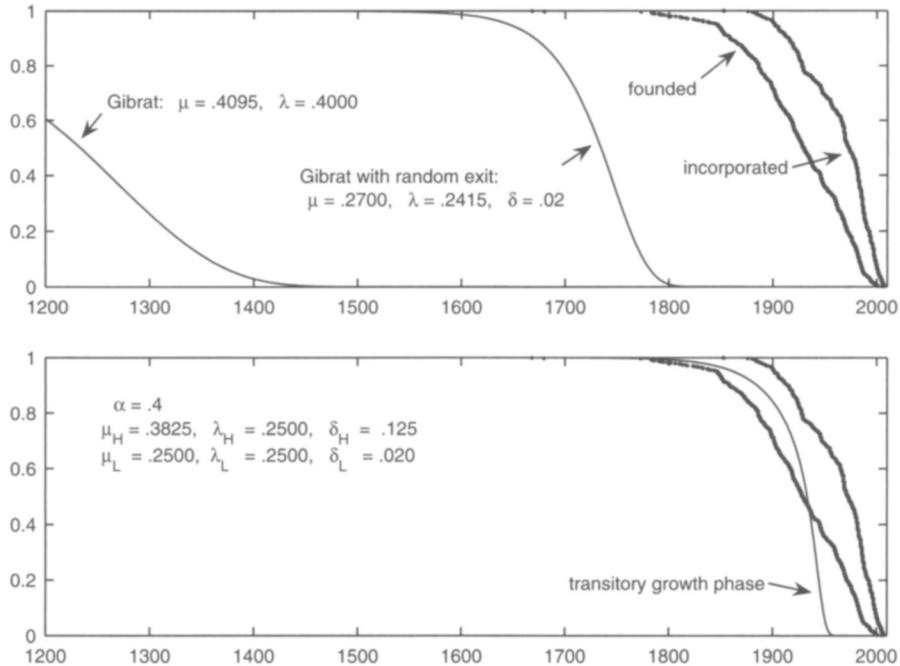


FIGURE 3
Distributions of firm startup years

and age distributions for a benchmark calibration in which a single employee is assigned to every blueprint. A fraction $\alpha = 0.4$ of new firms start out as high-growth firms. In the low-growth regime, the rates at which firms gain and lose blueprints are $\mu_L = \lambda_L = 0.2500$, and firms in this regime exit randomly at a rate $\delta_L = 0.02$.¹³ In the high-growth regime, firms also lose blueprints at the rate $\lambda_H = 0.2500$, but they gain blueprints at the higher rate $\mu_H = 0.3825$. Firms transition from the high-growth to the low-growth regime at a rate $\delta_H = 0.125$, resulting in a tail index $\zeta = (\eta + \delta_H)/(\mu_H - \lambda_H) \approx 1.02$. These parameters imply an exit rate of about 11% and Figures 2 and 3 show that these parameters closely match the observed size and age distributions. Holding fixed the other parameters, increasing the fraction α of high-growth new firms lowers the entry rate below and increases the number of large firms above what is observed in the data. Jointly changing (μ_H, δ_H) to $(\mu_H + \Delta, \delta_H + \Delta)$ changes the exit rate and the size distribution very little but shifts the age distribution of large firms. At $\delta_H = \delta_L = 0.02$, the age distribution is essentially that of the one-regime economy, while the median age of large firms can be as low as 50 years when $\delta_H = 0.25$.

But what about Gibrat's law? Many researchers find that Gibrat's law is a good approximation for firms that are not too small (*e.g.* Hall, 1987; Evans, 1987). Figure 4 shows the mean and standard deviation of the growth rates of surviving firms conditional on employment size. The mean growth rates are shown for three horizons: instantaneous, 1 year and 5 years. Survivorship bias affects the instantaneous mean only for firms with one employee. The probability of the

13. The U.S. Bureau of Labor Statistics reports monthly job separation rates for the 2000's that add up to around 50% on an annual basis. The parameter values $\lambda_H = \lambda_L = 0.25$ can be interpreted to mean that half of these separations do not correspond to job destruction, but to worker replacement.

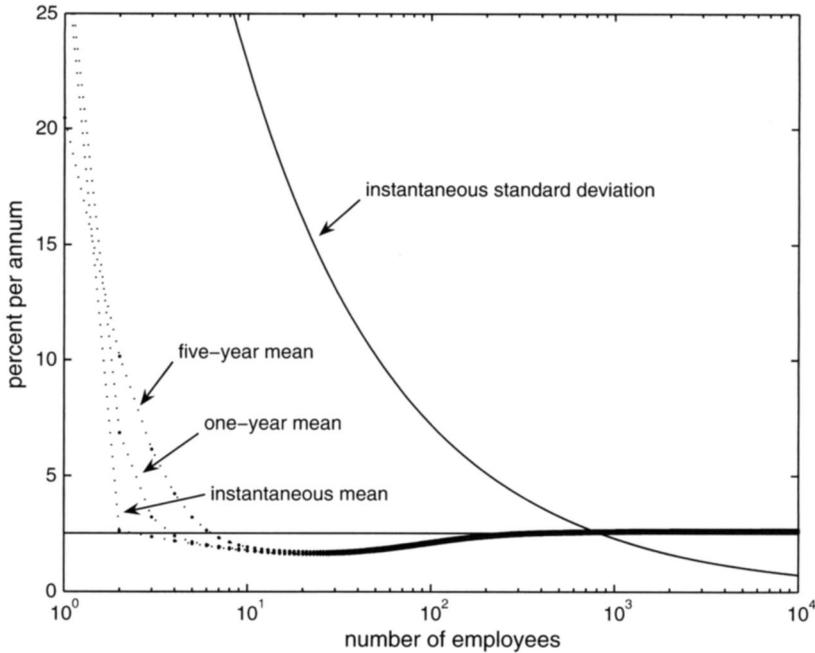


FIGURE 4
Growth rates by firm size

multiple blueprint losses it takes to induce exit is second order for firms with more than one employee. For the longer horizons commonly used in empirical studies this is no longer true, and surviving small firms grow much faster on average than the unconditional mean. But this effect declines rather quickly with size, and mean growth rates do not show much variation with size overall (the horizontal scale is logarithmic).

Over short intervals of time, the variance of surviving firm growth rates in phase I $\in \{H, L\}$ is $(\mu_1 + \lambda_1)/n$ for firm with $n > 1$ blueprints and μ_1 for firms with one blueprint. The calibration implies that this adds up to a standard deviation among all surviving firms of about 41% per annum. This is well within the range of standard deviations reported in Davis *et al.* (2007). For firms with more than one blueprint, the calibration implies a standard deviation of a firm with n blueprints is about $0.80/\sqrt{n}$ in the high-growth phase and $0.71/\sqrt{n}$ in the low-growth phase. This makes for very volatile growth rates among small firms. But for firms with more than 10 000 employees, these standard deviations will be less than 1% per annum. Even for a model without aggregate shocks, this is probably too small.

As emphasized by Klette and Kortum (2004), the empirical evidence suggests that the variance of firm growth rates declines more slowly than $1/n$. Hymer and Pashigian (1962) compared standard deviations of firm growth rates across size quartiles and found that firms in the largest quartile were significantly more volatile than predicted by the $1/n$ rule. More recently, Stanley *et al.* (1996) and Sutton (2002) find that the variance of the growth rate of Compustat firms behaves like $1/n^{1/3}$, and tentative interpretations are given in Stanley *et al.* (1996) and Sutton (2002, 2007). However, most firms are not publicly traded and are not covered by Compustat, and it is possible that these studies miss a rapid decline in variance that happens for small n .

5. CONCLUSION

Skewed firm size distributions are interpreted as reflecting skewed productivity distributions in Hopenhayn (1992), Atkeson and Kehoe (2005) and Luttmer (2007), among many others. The current paper attributes size differences not only to productivity differences but also to stochastic variation in the number of markets in which a firm operates, as in Klette and Kortum (2004), Lentz and Mortensen (2006) and Arkolakis (2006). Bounded productivity differences may give rise to unbounded size differences. In Lucas (1978), all variation in firm size is determined by heterogeneity in managerial talent. In Holmes and Schmitz (1995), Gabaix and Landier (2008) and Terviö (2008), both firm-specific productivity and managerial productivity play a role. Much remains to be done to sort out the relative importance of each of these aspects of firm heterogeneity.

Figure 1 and the relative young age of large firms are interpreted here using a two-phase pattern of growth in which some new firms start out with a high-quality blueprint and become firms with all low-quality blueprints after some random time. This is an abstraction that helps to illustrate the type of growth mechanism that can explain the size and age distribution of large firms. One expects more gradual declines in relative quality to work as well. A natural extension would allow for start-up blueprints that are initially of uncertain quality. This would bring in the selection considerations emphasized by Jovanovic (1982).

If blueprints are location specific, and locations are known to differ in how profitable they can be, then firms with new ideas will initially implement these in the more profitable locations, and only then expand, at a slower pace, into less attractive locations. This could be an alternative interpretation of the growth patterns shown in Figure 1, although it remains to be seen how this can account for the observed size distribution. One possibility is suggested by static models of Pareto-like size distributions. A well-known example is the Beckmann (1958) model of hierarchies of cities. More recently, Hsu (2007) describes an equilibrium model of hierarchies of firms and cities that produces Zipf's law. These static models could be viewed as long-run equilibrium conditions for a dynamic economy, and then the rapid initial growth shown in Figure 1 would simply reflect the fact that setting up a large firm is not quite instantaneous but still very fast.

Firms can grow along many margins. They can introduce new goods, build new plants, open new sales offices, hire new workers, win new customers, acquire whole new divisions. Entry sizes and the increments by which firms grow can vary across markets and industries. The framework sketched in this paper can be extended to incorporate these elements and arrive at a richer description of firm growth and heterogeneity. A close examination of the early histories of large U.S. corporations, such as those shown in Figure 1 and the ones described in Appendix A, shows that mergers, acquisitions and spin-offs are by no means infrequent. Along the lines of Jovanovic and Rousseau (2002), it is possible to interpret a small acquisition as the production of a new blueprint, but other interpretations are perhaps more natural. Spin-offs can give rise to firms that enter with a relatively large initial size, instead of the common minimum size assumed in this paper. It would be interesting to know if an account can be given of these aspects of firm growth that is consistent with the observed age and size distributions.

APPENDIX A

A.1. Firm employment and age data

The employment histories shown in Figure 1 were collected from Compustat, historical Moody's Manuals and corporate Web sites. Most employment histories are incomplete. Abbott Laboratories, Dow Chemical, IBM and Procter and Gamble were founded in the 19th century, and all other companies shown in Figure 1 during the 20th century.

The firm age data used in Section 4 were collected from several sources. Large firms are taken to be all Compustat firms headquartered in the U.S. with more than 10 000 employees, and firms in the same size category that appear on

a list of large privately held U.S. companies published by Forbes magazine, both in 2008. Compustat and Forbes tend to use a broad measure of firm employment. Employment at foreign subsidiaries is included, and franchisee employees appear to be counted as employees of the franchisor. The number of large firms obtained in this way is 813, which is somewhat less than the 953 firms with more than 10 000 employees reported by the SBA for 2006. In part, this may be because firms headquartered outside the U.S. are not included here. For example, Shell Oil Company is incorporated in the U.S. but is a wholly owned subsidiary of an English company headquartered in the Netherlands.

The date of incorporation of publicly traded firms was taken from the synopsis section of the Mergent Online database. For all firms, the foundation date is the date of the earliest reference to the company or its known predecessor companies that can be found in the company history section of the Mergent Online database or on company Web sites. In cases where this information is not available or does not appear to refer to the earliest times of the company, three additional sources were consulted: Dun and Bradstreet's Million Dollar Database, Hoover's Company Reports and the International Directory of Company Histories. In the case of privately held companies, Hoover's is the primary source. In a few cases, company age data were found in the Encyclopedia Britannica or in books available in the Google Books online library.

The Mergent Online database contains extensive records on now defunct corporations that were sometimes used to further trace back the origins of a company. Corporate Web sites of large companies often include extensive company histories that tend to emphasize the very old roots of the firm. Occasionally, the foundation date is taken to be the date its founding entrepreneur first started a business in the same industry, even if the company that eventually became large was not the first company started by the entrepreneur.

In the data collected here, mergers are an important source of firm growth that is not accounted for by the models in this paper. In many cases, a company history includes the year in which the oldest known component of a firm was founded. But in some industries, notably the health care industry, such information is almost non-existent. The models in this paper also do not allow for spinoffs. In the data set constructed here, a company that was already large at the time of its spinoff is taken to be founded at the time its parent was founded, when this information is available. Clearly, future empirical and theoretical work needs to account explicitly for mergers and spinoffs. The company age data together with the source for each age observation are available at www.luttmer.org.

APPENDIX B

B.1. Proof of Propositions 2, 3 and 4

Throughout this appendix, let $\theta, \mu \neq \lambda$ and $\theta - (\mu - \lambda)$ be positive, and take $\{A_n\}_{n=1}^\infty$ to be non-negative, not identically zero, and summable. Let $Y_0 = 0$ and consider the difference equation

$$\frac{1}{n}\theta Y_n = \lambda Y_{n+1} + \mu Y_{n-1} - (\lambda + \mu)Y_n + \frac{\mu A_n}{n} \tag{B.1}$$

for all $n \in \mathbb{N}$. Define $\beta_0 = 0$ and

$$\frac{1}{\beta_n} = 1 + \frac{\theta + \mu n}{\lambda n} - \frac{\mu \beta_{n-1}}{\lambda} \tag{B.2}$$

for all $n \in \mathbb{N}$. Then, the second-order difference equation (B.1) is equivalent to the pair of first-order difference equations

$$Y_n = Z_n + \beta_n Y_{n+1}, \quad Z_n = \frac{\mu \beta_n}{\lambda} \left[Z_{n-1} + \frac{A_n}{n} \right] \tag{B.3}$$

for all $n \in \mathbb{N}$, combined with the initial condition $Z_0 = 0$.

Lemma A1. *The sequence $\{\beta_n\}_{n=0}^\infty$ increases monotonically from 0 to $\min\{1, \lambda/\mu\}$. If $\mu > \lambda$, then $\lim_{n \rightarrow \infty} n(1 - \mu\beta_n/\lambda) = \theta/(\mu - \lambda)$. If $\mu < \lambda$, then $\lim_{n \rightarrow \infty} n(1 - \beta_n) = \theta/(\lambda - \mu)$.*

The proof can be given using diagrams for the recursions that define $\beta_n, n(1 - \mu\beta_{n-1}/\lambda)$ and $n(1 - \beta_{n-1})$, respectively. Note that $n = \infty$ in equation (B.2) yields a quadratic equation that is solved by $\beta_\infty \in \{1, \lambda/\mu\}$.

It is convenient to define two integrating factors,

$$B_n = \frac{1}{\beta_n} \prod_{k=1}^n \beta_k, \quad C_{n-1} = \left(\frac{\mu}{\lambda}\right)^{n-1} B_n$$

for all $n \in \mathbb{N}$. Note that $B_1 = C_0 = 1, C_n/C_{n-1} = \mu\beta_n/\lambda$, and equation (B.3) can be written as $B_n Y_n = B_n Z_n + B_{n+1} Y_{n+1}$ and $Z_n/C_n = Z_{n-1}/C_{n-1} + A_n/(nC_{n-1})$. We are interested only in solutions to equation (B.1) that are

non-negative and summable. Since $B_n \leq 1$, this implies that any such solution must satisfy $\lim_{n \rightarrow \infty} B_{n+1} Y_{n+1} = 0$. Given this boundary condition, equation (B.3) implies that

$$Y_n = \frac{1}{B_n} \sum_{k=0}^{\infty} B_{n+k} C_{n+k} \sum_{j=1}^{n+k} \frac{A_j}{j C_{j-1}} \quad (\text{B.4})$$

for all $n \in \mathbb{N}$. The following lemma collects some facts that are useful in determining the properties of Y_n .

Lemma A2. *Let $\{a_n\}_{n=1}^{\infty}$ be a positive sequence. If $\lim_{n \rightarrow \infty} a_{n+1}/a_n = \rho \in (0, 1)$, then*

$$\lim_{n \rightarrow \infty} n a_n \sum_{k=1}^n \frac{1}{k a_k} = \lim_{n \rightarrow \infty} \frac{1}{a_n} \sum_{k=0}^{\infty} a_{n+k} = \frac{1}{1-\rho}. \quad (\text{B.5})$$

Alternatively, if $\lim_{n \rightarrow \infty} n(1 - a_{n+1}/a_n) = \delta \in (0, \infty)$, then

$$\lim_{n \rightarrow \infty} a_n \sum_{k=1}^n \frac{1}{k a_k} = \lim_{n \rightarrow \infty} \frac{1}{a_n} \sum_{k=0}^{\infty} \frac{a_{n+k}}{n+k} = \frac{1}{\delta} \quad (\text{B.6})$$

and Raabe's test says that $\{a_n\}_{n=1}^{\infty}$ is summable if $\delta > 1$ and not if $\delta < 1$.

A version of equation (B.6) for regularly varying functions is discussed as part of Karamata's theorem in Bingham, Goldie and Teugels (1987). See Bojanic and Seneta (1973) for regularly varying sequences.

Suppose now that all A_n are positive and that $\lim_{n \rightarrow \infty} n(1 - A_{n+1}/A_n) = \zeta_A \in (0, \infty]$ is well defined. Also write $\gamma = \lim_{n \rightarrow \infty} A_{n+1}/A_n$ in situations in which this limit is well defined. Define $\zeta_C = \lim_{n \rightarrow \infty} n(1 - C_{n+1}/C_n)$. By Lemma A1, $\zeta_C = \theta/(\mu - \lambda) > 1$ if $\mu > \lambda$ and $\zeta_C = \infty$ if $\mu < \lambda$. The large- n behaviour of Y_n can be inferred from the fact that equation (B.4) can be represented as follows:

$$Y_n = T_n \sum_{k=0}^{\infty} w_{n,k} x_{n+k}, \quad (\text{B.7})$$

where T_n equals either A_n or C_n , and

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} w_{n,k} = \omega, \quad \lim_{n \rightarrow \infty} x_n = \zeta \quad (\text{B.8})$$

for some ω and ζ in $[0, \infty)$. This implies that $\lim_{n \rightarrow \infty} Y_n/T_n = \omega \zeta$.

When C_n declines more slowly than A_n , the following representation works

$$T_n = C_n, \quad w_{n,k} = \frac{B_{n+k} C_{n+k}}{B_n C_n}, \quad x_n = \sum_{j=1}^n \frac{A_j}{j C_{j-1}}. \quad (\text{B.9})$$

As long as $\mu \neq \lambda$, Lemma A1 implies $B_{n+1} C_{n+1}/(B_n C_n) \rightarrow \min\{\lambda/\mu, \mu/\lambda\} \in (0, 1)$ and then equation (B.5) gives $\omega = 1/(1 - \min\{\lambda/\mu, \mu/\lambda\})$. If $\mu/\lambda > 1$ and $\zeta_C < \zeta_A \leq \infty$, then Raabe's test implies that x_n converges to some $\zeta < \infty$. The same is true if $\gamma < \mu/\lambda < 1$.

When $\mu > \lambda$ and A_n declines more slowly than C_n , consider the representation

$$T_n = A_n, \quad w_{n,k} = \frac{A_{n+k} B_{n+k}}{A_n B_n}, \quad x_n = \left(\frac{A_n}{C_n}\right)^{-1} \sum_{j=1}^n \frac{A_j}{j C_{j-1}}.$$

Now Lemma A1 implies $A_{n+1} B_{n+1}/(A_n B_n) \rightarrow \lambda/\mu \in (0, 1)$ and equation (B.5) gives $\omega = 1/(1 - \lambda/\mu)$. Since $\zeta_A < \zeta_C$, (B.6) implies $\zeta = 1/(\zeta_C - \zeta_A)$.

When $\mu < \lambda$ and A_n declines more slowly than C_n , consider

$$T_n = A_n, \quad w_{n,k} = \frac{1}{n+k} \frac{A_{n+k} B_{n+k}}{A_n B_n}, \quad x_n = \frac{n C_n}{A_n} \sum_{j=1}^n \frac{A_j}{j C_{j-1}}.$$

Lemma A1 implies $n(1 - B_{n+1}/B_n) \rightarrow \theta/(\lambda - \mu) > 0$ and $C_n/C_{n-1} \rightarrow \mu/\lambda \in (0, 1)$. Furthermore, $n(1 - A_{n+1} B_{n+1}/(A_n B_n)) \rightarrow \zeta_A + \theta/(\lambda - \mu) > 0$ and thus (B.6) implies $\omega = 1/[\zeta_A + \theta/(\lambda - \mu)]$. If $\zeta_A < \infty$ then $(C_{n+1}/A_{n+1})/(C_n/A_n) \rightarrow \mu/\lambda \in (0, 1)$. Alternatively, if $\zeta_A = \infty$ and $\mu/\lambda < \gamma < 1$, then $(C_{n+1}/A_{n+1})/(C_n/A_n) \rightarrow [\mu/\lambda]/\gamma \in (0, 1)$. In either case, equation (B.5) implies that x_n converges to some $\zeta < \infty$.

Finally, note that the representation (B.7)–(B.9) also works and $\lim_{n \rightarrow \infty} Y_n/C_n \in (0, \infty)$ if $A_n > 0$ for only finitely many n .

Proposition A1. Consider the cases (i) $A_n > 0$ for finitely many n , (ii) $\lim_{n \rightarrow \infty} A_{n+1}/A_n = \gamma$ and (iii) $\lim_{n \rightarrow \infty} n(1 - A_{n+1}/A_n) = \zeta_A \in (0, \infty]$. Then, $\lim_{n \rightarrow \infty} Y_n/C_n \in (0, \infty)$ in case (i), in case (ii) if $1 > \mu/\lambda > \gamma$ and in case (iii) if $\zeta_A > \zeta_C$. Alternatively, $\lim_{n \rightarrow \infty} Y_n/A_n = 0$ in case (ii) if $1 > \gamma > \mu/\lambda$, and $\lim_{n \rightarrow \infty} Y_n/A_n \in (0, \infty)$ in case (iii) if $\zeta_A < \zeta_C$.

If we add the restriction $\zeta_A > 1$ when $A_n > 0$ for all n , then Raabe’s test implies that the solution (B.4) is in fact summable in all cases. One can generate many more solutions by adding a positive constant to $B_n Y_n$, but these cannot be summable.

Proposition A2. Suppose $A_n > 0$ for finitely many n , or $\lim_{n \rightarrow \infty} n(1 - A_{n+1}/A_n) = \zeta_A > 1$. Then,

$$\sum_{n=1}^{\infty} Y_n = \frac{1}{\theta - (\mu - \lambda)} \sum_{n=1}^{\infty} \mu A_n, \quad \sum_{n=1}^{\infty} \frac{1}{n} Y_n = \frac{1}{\theta} \left(\sum_{n=1}^{\infty} \frac{\mu A_n}{n} - \lambda Y_1 \right).$$

Proof. It follows from $\min\{\zeta_A, \zeta_C\} > 1$ that $nY_n \rightarrow 0$. Recall $Y_0 = 0$ and write equation (B.1) as follows:

$$\theta Y_n = \lambda[(n + 1)Y_{n+1} - nY_n - Y_{n+1}] - \mu[nY_n - (n - 1)Y_{n-1} - Y_{n-1}] + \mu A_n$$

for all $n \in \mathbb{N}$. Adding up over all n and using $\lim_{n \rightarrow \infty} nY_n = 0$ gives the first sum. The second sum follows from directly summing equation (B.1). \parallel

With these results in place, Proposition 2 follows from taking $\theta = \eta$, $A_1 = (\eta - (\mu - \lambda))/\mu$ and $A_{n+1} = 0$ for $n \in \mathbb{N}$. Proposition 3 then follows since part (B.6) of Lemma A2 shows that $\sum_{k=n}^{\infty} Y_k/k$ behaves like $Y_n \sim \omega \zeta C_n$, and $\zeta_C \in (1, \infty)$.

To prove Proposition 4, let $Q_{T,n}$ be the fraction of all blueprints in the economy held by type-T firms with n blueprints, where $T \in \{H, L\}$ and $n \in \mathbb{N}$. The properties of $Q_{H,n}$ are an application of Propositions 2 and 3. Proposition 4 can be shown by taking $\theta = \eta + \delta_L$, $(\mu, \lambda) = (\mu_L, \lambda_L)$, $A_1 = [(1 - \alpha)v + \delta_H Q_{H,1}]/\mu_L$, and $A_{n+1} = \delta_H Q_{H,n+1}/\mu_L$ for all $n \in \mathbb{N}$.

APPENDIX C

C.1. Proof of Proposition 5

C.1.1. Preliminaries. Suppose $\{X_i, Y_i\}_{i=1}^k$ are $2k$ independent random variables with $\Pr[X_i = n] = (1 - \gamma)\gamma^{n-1}$, $n \in \mathbb{N}$, $\Pr[Y_i = 0] = \theta$, and $\Pr[Y_i = 1] = 1 - \theta$. Define $Z_k = \sum_{i=1}^k X_i Y_i$ and let $K_k = \sum_{i=1}^k Y_i$.

As can be verified using moment generating functions, the sum of i.i.d. geometrically distributed random variables has a negative binomial distribution, given by

$$\Pr \left[\sum_{i=1}^m X_i = n \right] = \binom{n-1}{m-1} (1-\gamma)^m \gamma^{n-m}$$

for all $m \in \mathbb{N}$ and $n + 1 - m \in \mathbb{N}$. In view of the independence assumptions,

$$\Pr[Z_k = n] = \Pr \left[\sum_{i=1}^k X_i Y_i = n \right] = \sum_{m=1}^{\min\{k, n\}} \Pr \left[\sum_{i=1}^m X_i = n \right] \Pr[K_k = m]$$

for all $n \in \mathbb{N}$. Using the binomial distribution of K_k , this implies

$$\Pr[Z_k = n] = \sum_{m=1}^{\min\{k, n\}} \binom{k}{m} \binom{n-1}{m-1} (1-\theta)^m \theta^{k-m} (1-\gamma)^m \gamma^{n-m} \tag{C.1}$$

for all $n \in \mathbb{N}$. The complementary probability is $\Pr[Z_k = 0] = \theta^k$ since $Z_k = 0$ if and only if all Y_i are zero. Also, $\sum_{n=1}^{\infty} n \Pr[Z_k = n] = k \times (1 - \theta)/(1 - \gamma)$. This can be used to compute the mean growth rates conditional on survival reported in Figure 4.

Now suppose that K is drawn from the geometric distribution $(1 - \sigma)\sigma^{k-1}$, $k \in \mathbb{N}$. Then, the distribution of Z_K is determined by

$$(1 - \sigma) \sum_{k=1}^{\infty} \sigma^{k-1} \Pr[Z_k = n] = \frac{(1 - \sigma)\gamma^n}{\sigma(1 - \theta\sigma)} \sum_{m=1}^n \binom{n-1}{m-1} \left[\frac{\sigma(1 - \theta)(1 - \gamma)}{(1 - \theta\sigma)\gamma} \right]^m \quad (\text{C.2})$$

for all $n \in \mathbb{N}$. The right tail probabilities of this distribution are

$$(1 - \sigma) \sum_{n=N}^{\infty} \sum_{k=1}^{\infty} \sigma^{k-1} \Pr[Z_k = n] = \frac{1 - \theta}{1 - \theta\sigma} \left[\frac{(1 - \sigma)\gamma + \sigma(1 - \theta)}{1 - \theta\sigma} \right]^{N-1} \quad (\text{C.3})$$

for all $N \in \mathbb{N}$. For $N = 1$ this yields $\Pr[Z_k = 0] = (1 - \sigma)\theta/(1 - \sigma\theta)$.

Sketch of proof and computation. Suppose $\delta = 0$. Consider a firm that starts out with one blueprint. As reported in Klette and Kortum (2004), by age a such a firm will have exited with probability $\mathcal{T}_{0,1}(a) = \frac{\delta}{\mu} \gamma(a)$. Conditional on survival, its size distribution is the geometric size distribution $\mathcal{T}_{n,1}(a)/[1 - \mathcal{T}_{0,1}(a)] = [1 - \gamma(a)]\gamma^{n-1}(a)$. This can be verified directly by checking equations (26)–(27). The size distribution at age a of a firm that starts out with k blueprints is simply the distribution of the aggregate of k independent firms that start with one blueprint. Applying equation (C.1) gives $\{\mathcal{T}_{n,k}(a)\}_{n=1}^{\infty}$ for the case $\delta = 0$. Now suppose $\delta > 0$. Transitions from the first to the second phase occur at a rate δ , as long as no exit has taken place. This means that only a fraction $e^{-\delta a}$ of surviving firms remain in the initial phase. This determines $\{\mathcal{T}_{n,k}(a)\}_{n=1}^{\infty}$. The formulas for $\mathcal{T}_{-1,k}(a)$ and $\mathcal{T}_{0,k}(a)$ then follow from integrating equations (25)–(26), as described in the text.

The infinite sums needed in equations (28) and (29) follow from equations (C.2) and (C.3). Age densities (distributions) can then be computed using a univariate (bivariate) numerical integration.

Acknowledgment. This paper has evolved from my Federal Reserve Bank of Minneapolis Working Papers 645, 649 and 657. The data and more detailed proofs are available at www.luttmer.org. I thank Nathalie Pouokam for skillful research assistance and two anonymous referees for helpful comments. 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.

REFERENCES

- ARKOLAKIS, C. (2006), "Market Access Costs and the New Consumers Margin in International Trade" (Working Paper, University of Minnesota).
- ARMINGTON, C. (1998), *Statistics on U.S. Businesses—Microdata and Tables* (US Small Business Administration).
- ATKESON, A. and KEHOE, P. (2005), "Modelling and Measuring Organization Capital", *Journal of Political Economy*, **113**, 1026–1051.
- AXTELL, R. L. (2001), "Zipf Distribution of U.S. Firm Sizes", *Science*, CCXCIII, 1818–1820.
- BECKMANN, M. J., (1958), "City Hierarchies and the Distribution of City Size", *Economic Development and Cultural Change*, **6**, 243–248.
- BERNARD, A. B., REDDING, S. J. and SCHOTT, P. K. (2006), "Multi-Product Firms and Product Switching" (NBER Working Paper No. 12293).
- BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987), *Regular Variation* (Cambridge: Cambridge University Press).
- BOJANIC, R. and SENETA, E. (1973), "A Unified Theory of Regularly Varying Sequences", *Mathematische Zeitschrift*, **134**, 91–106.
- BOLDRIN, M. and LEVINE, D. K. (1999, 2006), "Perfectly Competitive Innovation" (Working Paper, Washington University).
- CHATTERJEE, S. and ROSSI-HANSBERG, E. (2006), *Spin-offs and the Market for Ideas* (Working Paper, Princeton University).
- DAVIS, S. J., HALTIWANGER, J., JARMIN, R. and MIRANDA, J. (2007), "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms", in Acemoglu, D., Rogoff, K. and Woodford, M. (eds) *NBER Macroeconomics Annual 2006*, Vol. 21 (Cambridge: MIT Press), 107–156.
- DIXIT, A. K. and STIGLITZ, J. E. (1977), "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, **LXVII**, 297–308.
- DUNNE, T., ROBERTS, M. J. and SAMUELSON, L. (1989), "The Growth and Failure of U.S. Manufacturing Plants", *Quarterly Journal of Economics*, **CIV**, 671–698.
- EVANS, D. S. (1987), "The Relationship between Firm Growth, Size and Age: Estimates for 100 Manufacturing Industries", *Journal of Industrial Economics*, **XXXV**, 567–581.
- GABAIX, X. (1999), "Zipf's Law for Cities: An Explanation", *Quarterly Journal of Economics*, **CXIV**, 739–367.

- GABAIX, X. and LANDIER, A. (2008), "Why Has CEO Pay Increased so Much", *Quarterly Journal of Economics*, **123**, 49–100.
- GROSSMAN, G. and HELPMAN, E. (1991), *Innovation and Growth in the Global Economy* (Cambridge: MIT Press).
- HALL, B. H., (1987), "The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector", *Journal of Industrial Economics*, **XXXV**, 583–605.
- HOLMES, T. J. (2006), "The Diffusion of Wal-Mart and Economies of Density" (Working Paper, University of Minnesota).
- HOLMES, T. J. and SCHMITZ, J. A., Jr. (1995), "On the Turnover of Business Firms and Business Managers", *Journal of Political Economy*, **103**, 1005–1038.
- HOPENHAYN, H. (1992), "Entry, Exit and Firm Dynamics in Long Run Equilibrium", *Econometrica*, **60**, 1127–1150.
- HSU, W.-T. (2007), "Central Place Theory and Zipf's Law" (Working Paper, University of Minnesota).
- HYMER, S. and PASHIGIAN, P. (1962), "Firm Size and Rate of Growth", *Journal of Political Economy*, **70**, 556–569.
- IJIRI, Y. and SIMON, H. A. (1964), "Business Firm Growth and Size", *American Economic Review*, **LIV**, 77–89.
- JACKSON, M. O. (2006), "The Economics of Social Networks", in Blundell, R., Newey, W. and Persson, T. (eds) *Proceedings of the 9th World Congress of the Econometric Society* (Cambridge: Cambridge University Press) 1–56.
- JACKSON, M. O. and ROGERS, B. W. (2007), "Meeting Strangers and Friends of Friends: How Random Are Social Networks", *American Economic Review*, **97**, 890–915.
- JOVANOVIC, B. (1982), "Selection and the Evolution of Industry", *Econometrica*, **C**, 649–670.
- JOVANOVIC, B. and ROUSSEAU, P. L. (2002), "A Q-Theory of Mergers", *American Economic Review, Papers and Proceedings*, **92**, 198–204.
- KENDALL, D. G. (1948), "On the Generalized Birth-and-Death Process", *The Annals of Mathematical Statistics*, **19**, 1–15.
- KLETTE, T. J. and KORTUM, S. S. (2004), "Innovating Firms and Aggregate Innovation", *Journal of Political Economy*, **112**, 986–1018.
- KRUGMAN, P. (1996), *The Self-Organizing Economy* (Oxford: Blackwell Publishers).
- LENTZ, R. and MORTENSEN, D. T. (2006), "An Empirical Model of Growth Through Product Innovation" (University of Wisconsin and Northwestern University).
- LUCAS, R. E., Jr. (1978), "On the Size Distribution of Business Firms", *Bell Journal of Economics*, **9**, 508–523.
- LUTTMER, E. G. J. (2006), "Consumer Search and Firm Growth" (Working Paper No. 645, Federal Reserve Bank of Minneapolis).
- LUTTMER, E. G. J. (2007), "Selection, Growth, and the Size Distribution of Firms", *Quarterly Journal of Economics*, **122**, 1103–1144.
- LUTTMER, E. G. J. (2008), "On the Mechanics of Firm Growth" (Working Paper No. 657, Federal Reserve Bank of Minneapolis).
- PRESCOTT, E. C. and VISSCHER, M. (1980), "Organization Capital", *Journal of Political Economy*, **88**, 446–461.
- ROMER, P. M. (1990), "Endogenous Technical Change", *Journal of Political Economy*, **98**, S71–S102.
- ROSEN, S. (1978), "Substitution and Division of Labour", *Economica*, **45**, 235–250.
- ROSSI-HANSBERG, E. and WRIGHT, M. L. J. (2007), "Establishment Size Dynamics in the Aggregate Economy", *American Economic Review*, **97**, 1639–1666.
- SEKER, M. (2007), "A Structural Model of Establishment and Industry Evolution: Evidence from Chile" (Working Paper, University of Minnesota).
- SIMON, H. A. (1955), "On a Class of Skew Distribution Functions", *Biometrika*, **42**, 425–440.
- SIMON, H. A. and BONINI, C. P. (1958), "The Size Distribution of Business Firms", *American Economic Review*, **XCVIII**, 607–617.
- STANLEY, R. M., AMARAL, L. A. N., BULDYREV, S. V., HARLIN, S., LESCHORN, H., MAASS, P., SALINGER, M. A., STANLEY, H. E. (1996), "Scaling Behavior in the Growth of Companies", *Nature*, **319**, 804–806.
- SUTTON, J. (1997), "Gibrat's Legacy", *Journal of Economic Literature*, **XXXV**, 40–59.
- SUTTON, J. (2002), "The Variance of Firm Growth Rates: The 'Scaling' Puzzle", *Physica A*, **CCCXII**, 577–590.
- SUTTON, J. (2007), "Market Share Dynamics and the 'Persistence of Leadership' Debate", *American Economic Review*, **97**, 222–241.
- TERVIÖ, M. (2008), "The Difference that CEOs Make: An Assignment Model Approach", *American Economic Review*, **98**, 642–668.
- YOUNG, A. (1998), "Growth Without Scale Effects", *Journal of Political Economy*, **106**, 41–63.
- YULE, G. U. (1925), "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*, **213**, 21–87.