

On the Mechanics of Firm Growth¹

Erzo G.J. Luttmer
University of Minnesota
and
Federal Reserve Bank of Minneapolis

October 6, 2009

¹An earlier version of this paper circulated under the title “New Goods and the Size Distribution of Firms.” I have received useful comments and suggestions from seminar participants at UCLA, the Federal Reserve Bank of Minneapolis, and the SED and AEA meetings in Prague and New Orleans. 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.

Abstract

Given a common technology for replicating blueprints, high-quality blueprints will be replicated more quickly than low-quality blueprints. If trade in blueprints is subject to small transaction costs, then firms can be identified with collections of blueprints derived from the same initial blueprint. Along the balanced growth path, Gibrat's Law holds conditional on firm type. A stationary firm size distribution with the thick right tail observed in the data can then arise only when the number of blueprints in the economy grows over time, or else firms cannot grow at a positive rate on average. But when calibrated to match the observed firm entry rate and the right tail of the size distribution, this model implies that the median age among firms with more than 10,000 employees is about 750 years. The problem is Gibrat's Law. If a firm with high-quality blueprints can lose its quality lead, then such a firm's growth rate will slow down eventually. By allowing for rapid and noisy initial growth, this version of the model can explain high observed entry rates, a thick-tailed size distribution, and the relatively young age of large U.S. corporations.

1. INTRODUCTION

Why does the employment size distribution of firms look like a Pareto distribution with $\Pr[\text{number of firm employees} \geq n] \sim n^{-\zeta}$ and a tail index $\zeta \approx 1.05$, barely large enough for the distribution to have a finite mean? Why are there so many large firms and how did they grow so large?

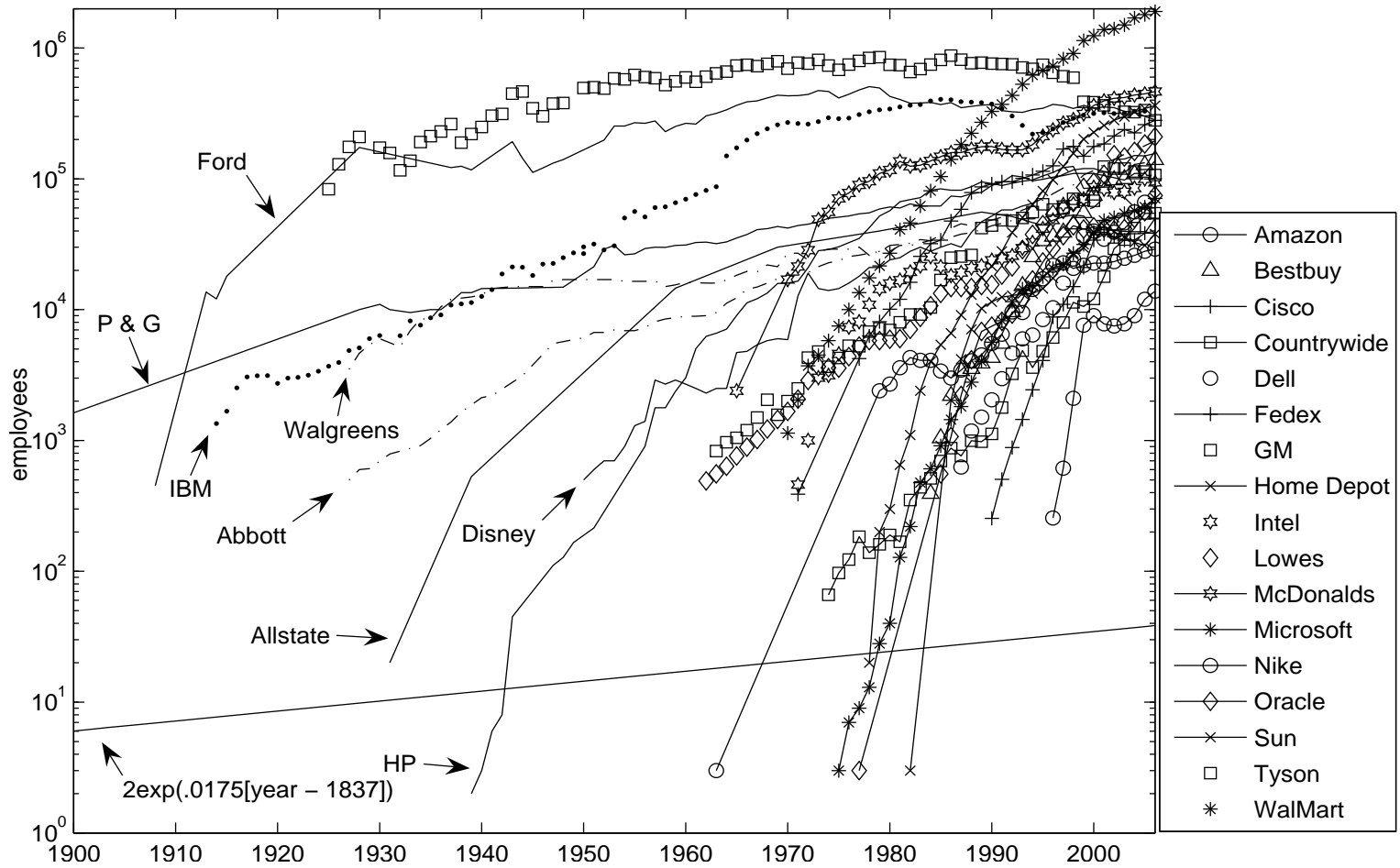
In the presence of decreasing returns or downward sloping firm demand curves, it is possible that the highly skewed size distribution entirely reflects a highly skewed productivity distribution. Such a productivity distribution can arise from random firm-level productivity growth, combined with selection of relatively productive firms. Random growth implies a strict version of Gibrat's law: firm growth rates are independent of size. This is widely thought to be a reasonable first approximation to the data. If incumbent firms become more productive at an average rate that does not exceed the productivity growth rate of new entrants, and unproductive firms are driven out of business, then a stationary size distribution will result. This distribution will have a tail index ζ just above 1 when there is only a small gap between entrant and incumbent mean productivity growth rates (Luttmer [2007]).¹

But is size really just productivity, and are mean growth rates really constant? At the very least, it is likely to take time for a firm's technological advances to be reflected in its size. Figure 1 presents some striking evidence on firm growth rates. It shows the employment histories of 25 large firms with more than 10,000 employees in 2004. The U.S. Census reports that there were close to a thousand such firms in 2004, and that firms in this size class accounted for slightly over a quarter of U.S. employment. The employment data reported in Figure 1 are from Compustat, filings with the Securities and Exchange Commission, and company web sites.² Included, for comparison, is an artificial non-stochastic growth path for the Procter & Gamble Company that would

¹The $\zeta = 1$ 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. Galbaix [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 [2006] 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.

²Davis et al. [2007] point out the difficulties associated with self-reported employment statistics such as these. Chandler [1994, p. 137] shows a pattern very similar to Figure 1 for gross fixed assets in the U.S. steel industry over the period 1900-1950.

Figure I



have resulted if employment at this company had grown, since its foundation in 1837, at the U.S. population growth rate. The mean employment growth rate across all firms reported in Figure I is a little over 18% per annum, and there is considerable variation. In particular, firm growth rates seem to be persistently much above average when firms are relatively small. And they decline significantly when firms become large. Most of the sample paths reported in this figure, taking into account that they are selected to generate large firms, are unlikely to have been generated by the geometric random walk that gives rise to Gibrat's law.³

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. Simon and Bonini [1958], Ijiri and Simon [1964], and many others since studied firm growth. To generate a Pareto-like size distribution, these papers rely on growth in the number of firms and a weak form of Gibrat's law: mean growth rates are independent of size. This is the starting point here as well, but then the Gibrat assumption is relaxed to account for the evidence shown in Figure I, in a way that continues to imply a stationary size distribution that matches the empirical size distribution of U.S. firms. An explicit formula is obtained for the tail index ζ in terms of the parameters of the growth process of firms.

In the model, firms produce differentiated commodities using labor and commodity-specific blueprints. An entrepreneur can set up a new firm by creating a blueprint of a certain quality. Once set up, the firm can use labor and any of its blueprints to create new blueprints of the same quality. Individual blueprints can also become obsolete. The arrival rates of these two events are independent and independent across blueprints. In the Gibrat version of the model, blueprints are of constant quality until the time they become obsolete. As a result, the mean growth rate of a firm with more than a single blueprint is independent of firm size. In this environment, firms with high-quality blueprints have stronger incentives to replicate their blueprints, and hence grow faster, than firms with low-quality blueprints.

Whether or not blueprints vary in quality, a setup like this can generate a firm size distribution with a thick right tail, provided that the aggregate number of blueprints grows at a strictly positive rate—equal to the population growth rate along the balanced growth path—and some types of firms gain blueprints at almost this rate. If the

³Besides Procter & Gamble, the only other firms in Figure I that were founded in the 19th century are Abbott (1888), GM (1897), and IBM (1889).

aggregate number of blueprints were constant, then the fact that there is entry would imply that incumbent firms must lose blueprints on average. When the mean growth rate of firms is constant, this rate must be negative, and few firms can become very large. The size distribution in Klette and Kortum [2004] has a thin right tail for precisely this reason.

Calibrating an economy with homogeneous blueprints using the tail index $\zeta \approx 1.05$ and the observed firm entry rate of about 10% per annum implies that the median firm with 10,000 or more employees is about 750 years old. In U.S. data, the median age of such firms is about 75 years. Although the Procter & Gamble Company was not built in a day, or even 75 years, it did not take 750 years. To capture this, it is assumed that a firm's blueprints can depreciate in quality without becoming completely obsolete. A new firm enters as a high-quality firm with a high-quality blueprint. After some random time, the quality of the firm and all of its blueprints reverts to "normal," or low quality. As a result, firms initially choose to grow faster than they will eventually, as appears to be the case in Figure I. Depending on the shape of the replication technology, these growth rate differences can be large even for small differences in blueprint quality. In contrast to Luttmer [2007], bounded differences in quality can lead to unbounded differences in firm size. Rapid initial growth is consistent with a stationary size distribution if the mean duration of the initial growth phase is not too long. A relatively young population of very large firms arises if firms tend to be large not because of long-term growth rates that are close to the population growth rate, but because of the very rapid initial growth of new firms. High observed entry rates imply that growth must be noisy, so that small firms may exit with a high probability. Importantly, the calibrated regression of firm growth rates on size is such that only small firms appear to violate Gibrat's law.

Entry decisions and blueprint replication rates are endogenous in this paper, depending ultimately on the talent distribution of entrepreneurs, the productivity of the blueprint replication technologies, and the extent to which the replication of any single blueprint is subject to decreasing returns. All agents can supply labor or develop new blueprints, but talented agents have a comparative advantage in producing new blueprints. If the entrepreneurial talent distribution has unbounded support, then some entry will take place at any combination of wages and blueprint prices, and the balanced growth path forces entry and replication rates to be such that the size distribution is stationary.⁴ For simple cases, one can examine the dynamics away from the balanced

⁴In Luttmer [2007], incumbent growth rates are exogenous but a spillover makes the entry and aggregate productivity growth rates endogenous. There, a weak spillover gives rise to a size distribution

growth path and confirm that the balanced growth path is stable. The speed of adjustment can be slow and depends on whether the number of blueprints is above or below the balanced growth path.

As already indicated, this paper solves the thin-right-tail problem in Klette and Kortum [2004]. It does so in an analytically tractable and inevitably stylized setup. Lentz and Mortensen [2007] 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.⁵ 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 Luttmner [2006]. Related models of network formation are presented in Jackson [2006] and Jackson and Rogers [2007], and the extensive literature cited therein.

The economy is introduced in Section 2 and its balanced growth path is described in Section 3, together with two alternative formulations of the role of blueprints in production. The stationary size distribution is derived in Section 4 and formulas are given for the tail index ζ in the Gibrat and non-Gibrat cases. Section 5 constructs the age distribution conditional on size. These results are used in Section 6 to interpret age and size data on U.S. firms. Section 7 concludes.

with a thick right tail.

⁵See also Seker [2007] for related work on Chilean establishments.

2. THE ECONOMY

Blueprints contain ideas and are costly to produce and replicate. In the baseline specification, a blueprint describes the idea for a final good. Replication generates a blueprint for a distinct final good. Final goods producers are monopolistic competitors. A competitive alternative in which blueprints specify plants that are subject to decreasing returns is discussed later. The equilibria in both economies are Pareto efficient.

2.1 Consumers

Time is continuous and indexed by $t \in [0, \infty)$. 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 sequences 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

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

Consumption is a composite good that consists of a continuum of differentiated commodities produced by producers who are of types taken from a countable set \mathcal{Z} . There is a measure $N_{z,t}$ of producers of type $z \in \mathcal{Z}$, and all these producers charge the same price $p_{z,t}$ in equilibrium. Aggregate consumption of a commodity trading at a price p is $C_t(p)$. The composite good is as in Dixit and Stiglitz [1977],

$$C_t = \left[\sum_{z \in \mathcal{Z}} C_t^\beta(p_{z,t}) N_{z,t} \right]^{1/\beta}, \quad (2)$$

with $\beta < 1$ restricted to also be positive so that consumers have a preference for variety. Consumers choose $C_t(\cdot)$ to minimize the cost of acquiring C_t . The resulting demand functions are

$$C_t(p) = \left(\frac{p}{P_t} \right)^{-1/(1-\beta)} C_t, \quad (3)$$

where P_t is the price index

$$P_t = \left[\sum_{z \in \mathcal{Z}} p_{z,t}^{-\beta/(1-\beta)} N_{z,t} \right]^{-(1-\beta)/\beta}. \quad (4)$$

Note that prices of differentiated commodities are quoted in some arbitrary numeraire. All other prices will be expressed in units of the composite commodity.

2.2 Producers

A producer charging a price p must supply $C_t(p)$ and this generates revenues equal to $C_t^{1-\beta}C_t^\beta(p)$, measured in units of composite consumption. The production of a differentiated commodity requires a blueprint. Given a type- z blueprint, a producer can use l units of labor to produce $zQ_t l$ units of its differentiated commodity. The labor-augmenting productivity component Q_t is common to all producers and evolves exogenously according to $Q_t = Q_t e^{\theta t}$. One possible interpretation is that Q_t is labor-augmenting human capital that agents choose to accumulate at a constant rate. Alternatively, the model can be expanded to include vintage equipment as a factor of production. Then Q_t would represent the average quality at time t of equipment used by commodity producers. In either case, a commodity producer's type z augments the productivity of inputs that are used by all producers.

At time t , a type- z producer chooses labor inputs l to maximize $C_t^{1-\beta}(zQ_t l)^\beta - w_t l$. The solution is

$$l_{z,t} = \left(\frac{\beta z Q_t}{w_t} \right)^{\beta/(1-\beta)} \frac{\beta C_t}{w_t}. \quad (5)$$

Measured in units of labor, this yields revenues $\psi l_{z,t}$, where $\psi = (1 - \beta)/\beta$. All type- z producers charge the price $p_{z,t}$ implied by (3) and supply

$$C_t(p_{z,t}) = z Q_t l_{z,t} \quad (6)$$

units of their differentiated commodity. Inserting this into (2) and solving for w_t gives

$$w_t = \beta Q_t \left[\sum_{z \in \mathcal{Z}} z^{\beta/(1-\beta)} N_{z,t} \right]^{(1-\beta)/\beta}. \quad (7)$$

Note that the elasticity of wages with respect to a proportional increase in all $N_{z,t}$ is ψ .

2.3 New Blueprints

A type- z producer needs a type- z blueprint to produce. Blueprints depreciate in one-hoss-shay fashion at an average rate $\lambda_{z,t}$. New blueprints can be created by using labor to replicate existing blueprints, or from scratch by entrepreneurs. The respective rates at which this occurs in equilibrium are denoted by $\mu_{z,t}$ and $\nu_{z,t}$. The number of new type- z blueprints therefore evolves according to

$$DN_{z,t} = (\nu_{z,t} + \mu_{z,t} - \lambda_{z,t}) N_{z,t}. \quad (8)$$

An initial condition determines $N_{z,0}$, for each $z \in \mathcal{Z}$.

2.3.1 Replication of Existing Blueprints

A new blueprint of type z can be created from an existing type- z blueprint. The new blueprint arrives following an exponentially distributed waiting time with mean $\mu_{z,t} = f(i_{z,t})$, where $i_{z,t}$ is labor employed in the replication process. An existing blueprint is lost following an exponentially distributed waiting time with mean $\lambda_{z,t} = g(j_{z,t})$, where $j_{z,t}$ is labor 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.⁶ The value $v_{z,t}$ of a type- z blueprint must satisfy the Bellman equation

$$r_t v_{z,t} = \max_{\substack{\mu \leq f(i) \\ \lambda \leq g(j)}} \{w_t (\psi l_{z,t} - [i + j]) + (\mu - \lambda) v_{z,t} + D_t v_{z,t}\}, \quad (9)$$

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, and $g(0)$ represents the rate at which blueprints are lost without any effort. For convenience, both f and g are assumed to sufficiently smooth with unbounded slopes near zero.⁷ The optimal investment in new blueprints is determined by

$$\mu_{z,t} = f(i_{z,t}), \quad \lambda_{z,t} = g(j_{z,t}), \quad v_{z,t} D_t f(i_{z,t}) = -v_{z,t} D_t g(j_{z,t}) = w_t. \quad (10)$$

The technology assumptions ensure that $\mu_{z,t}$ and $-\lambda_{z,t}$ are increasing in $v_{z,t}$. High-value blueprints are replicated more quickly and maintained better than low-value blueprints.

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. Not every new blueprint is of the most

⁶The 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 Luttmner [2007], spillovers are assumed to be of secondary importance in this economy.

⁷The replication and maintenance technology is the same for all types of blueprints. It is conceptually straightforward to allow for heterogeneity across types, and this could very well be an important source of variation in firm growth rates.

productive type because entrepreneurs weigh the value of high-quality blueprints against the cost of the time it takes to design them. At any point in time, every agent in the economy is endowed with one unit of effort that can be assigned to various tasks: supply labor, or attempt to produce a type- z blueprint for some $z \in \mathcal{Z}$. Every agent has a skill vector (x, y) , where $x = \{x_z\}_{z \in \mathcal{Z}}$ are the rates at which the agent can develop new blueprints of different types and y is the amount of labor the agent can supply per unit of time. Given wages w and a vector of blueprint prices $v = \{v_z\}_{z \in \mathcal{Z}}$, define

$$X_z[v, w] = \left\{ (x, y) : v_z x_z = \max_{z' \in \mathcal{Z}} \{v_{z'} x_{z'}\} \geq wy \right\}$$

for all $z \in \mathcal{Z}$, and

$$Y[v, w] = \left\{ (x, y) : wy \geq \max_{z' \in \mathcal{Z}} \{v_{z'} x_{z'}\} \right\}.$$

Comparative advantage determines occupational choice. Ignoring ties, agents with a skill vector in $X_z[v, w]$ choose to be entrepreneurs who design type- z blueprints, and agents with a skill vector in $Y[v, w]$ will choose to be employees. Given prices (v, w) , an agent with skill vector (x, y) earns $\max\{(v_z x_z)_{z \in \mathcal{Z}}, wy\}$ per unit of time.

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]. The resulting per-capita supply of type- z entrepreneurial effort is

$$E_z(v_t, w_t) = \int_{X_z[v_t, w_t]} x_z dT(x, y) \quad (11)$$

for $z \in \mathcal{Z}$. The per-capita supply of labor is

$$L(v_t, w_t) = \int_{Y[v_t, w_t]} y dT(x, y). \quad (12)$$

Clearly, $E_z(\cdot)$ and $L(\cdot)$ are both homogeneous of degree 1, and the supply of every activity is increasing in its own price—blueprint price or wage—and decreasing in all other prices.

2.4 Equilibrium

Given a per-capita supply of entrepreneurial effort $E_z(v_t, w_t)$ and a stock of type- z blueprints $N_{z,t}$, the rate $\nu_{z,t}$ at which entrepreneurs add new type- z blueprints is determined by

$$\nu_{z,t} N_{z,t} = H_t E_z(v_t, w_t) \quad (13)$$

for each $z \in \mathcal{Z}$. Labor market clearing requires that

$$\sum_{z \in \mathcal{Z}} N_{z,t} (l_{z,t} + i_{z,t} + j_{z,t}) = H_t L(v_t, w_t). \quad (14)$$

The equilibrium is determined by (1)-(14), initial conditions $\{N_{z,0}\}_{z \in \mathcal{Z}}$, and transversality conditions for $\{v_{z,t}, N_{z,t}\}_{z \in \mathcal{Z}}$.

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. If there is only one type of blueprint, then it is possible to characterize the dynamics in terms of only one state and one costate variable, and construct an equilibrium that converges over time to a balanced growth path.⁸

3. BALANCED GROWTH AND ALTERNATIVE FORMULATIONS

Along a balanced growth path, the measure of blueprints of every type grows at the rate η , and thus $N_{z,t} = N_z e^{\eta t}$ for some N_z . The allocation of labor per type- z blueprint is constant at (i_z, j_z, l_z) . Because of (2), (6) and (7), per capita consumption and wages grow at the rate $\kappa = \theta + \psi\eta$. The implied interest rate is $r = \rho + \gamma\kappa$. Using (5), production labor per type- z blueprint can be written as

$$l_z = z^{1/\psi} l, \quad z \in \mathcal{Z}, \quad (15)$$

for some positive l . The resulting revenues per blueprint are necessarily positive and hence blueprint prices will be positive. The Bellman equation (9) implies that wages $w_t = w e^{\kappa t}$ and blueprint prices $v_{z,t} = v_z e^{\kappa t}$ must satisfy the present-value condition

$$\frac{v_z}{w} = \frac{\psi l_z - [i_z + j_z]}{r - \kappa - [\mu_z - \lambda_z]}, \quad z \in \mathcal{Z}, \quad (16)$$

where (i_z, j_z) and (μ_z, λ_z) are determined by the optimality requirements

$$\begin{bmatrix} \mu_z \\ \lambda_z \end{bmatrix} = \begin{bmatrix} f(i_z) \\ g(j_z) \end{bmatrix}, \quad \frac{v_z}{w} \begin{bmatrix} Df(i_z) \\ -Dg(j_z) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad z \in \mathcal{Z}. \quad (17)$$

⁸The 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. It is possible to generate slow and asymmetric adjustment to the balanced growth path. A detailed analysis is beyond the scope of this paper.

Since positive revenues net of replication and maintenance costs are feasible, (16) must hold with $\psi l_z > i_z + j_z$ and $r - \kappa > \mu_z - \lambda_z$. The fact that the aggregate number of blueprints grows at the rate η implies that entrepreneurs must contribute new blueprints at the non-negative rate $\eta - [\mu_z - \lambda_z]$. If $E_z(v, w)$ is positive, then the entrepreneurial supply of blueprints (13) determines the number of blueprints via

$$\frac{N_z}{H} = \frac{E_z(v, w)}{\eta - [\mu_z - \lambda_z]}, \quad z \in \mathcal{Z}. \quad (18)$$

Alternatively, $E_z(v, w) = 0$ and $(\eta - [\mu_z - \lambda_z])N_z/H = 0$. Along a balanced growth path, the labor-market clearing condition (14) becomes

$$\sum_{z \in \mathcal{Z}} \frac{N_z}{H} (i_z + j_z + l_z) = L(v, w). \quad (19)$$

Given some positive scale factor l , the balanced growth conditions (15)-(17) determine the blueprint prices v_z/w , the labor allocations (i_z, j_z, l_z) , and the resulting accumulation rates μ_z and λ_z . The supply of blueprints (18) and the labor-market clearing condition (19) then pin down the level of l . Note that these equilibrium conditions only depend on blueprint prices relative to wages. The level of wages follows from (7) and (18). Aggregate consumption can be obtained from (5).

The present value of aggregate consumption must be finite, and this requires parameter values so that $r > \kappa + \eta$. Since the number of blueprints grows at the rate η along any balanced growth path, replication of existing blueprints can contribute at a rate of at most $\mu_z - \lambda_z \leq \eta$, and this will hold for all small enough i_z and j_z as long as $\eta > f(0) - g(0)$. Together, these inequalities imply $r - \kappa > \mu_z - \lambda_z$ for all $z \in \mathcal{Z}$, and this ensures finite present values. The optimality of i_z and j_z implies that $(r - \kappa)v_z/w \geq \psi l_z - [i_z + j_z] + [\mu_z - \lambda_z]v_z/w$ for any $z' \in \mathcal{Z}$, and thus

$$(r - \kappa - [\mu_{z'} - \lambda_{z'}])(v_z - v_{z'})/w \geq \psi(l_z - l_{z'})$$

for any two z and z' in \mathcal{Z} . This implies that v_z/w is increasing in z , since l_z is increasing in z , by (15). The assumption that f and $-g$ are concave then implies the same for μ_z and $-\lambda_z$. A similar argument implies that the blueprint prices v_z/w that satisfy (15)-(17) are increasing in l . In turn, the entrepreneurial supply of blueprints is increasing in blueprint prices. The left-hand side of (19), taking into account (15)-(18), is therefore increasing in l . In particular, the demand for labor grows without bound as l increases to a point where $\mu_z - \lambda_z$ approaches η for the most productive blueprint. The supply of labor on the right-hand side (19) is decreasing in l , again taking into account how

blueprint prices depend on l via (15)-(17). Together these observations can be used to prove the following.

PROPOSITION 1 *Suppose that $\rho + \gamma\kappa > \kappa + \eta$ and $\eta > f(0) - g(0)$. Take the set of blueprint types \mathcal{Z} to be finite and suppose that the talent distribution is such that $E_z(v, w) > 0$ for all $z \in \mathcal{Z}$ and all strictly positive (v, w) . Then (15)-(19) defines a unique balanced growth path that satisfies $\eta > \mu_z - \lambda_z$ for all $z \in \mathcal{Z}$. The employment levels i_z , j_z and l_z as well as the accumulation rates μ_z and λ_z are increasing in z .*

The assumption that \mathcal{Z} is finite is the simplest way to ensure that there is a most productive type of blueprint. Together with $E_z(v, w) > 0$ this then implies that the $\eta - [\mu_z - \lambda_z]$ are positive and bounded away from zero. An equilibrium with $E_z(v, w) = 0$, possibly for all $z \in \mathcal{Z}$, 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. Since $\eta - [\mu_z - \lambda_z]$ is decreasing in z , it must then be that $N_z > 0$ and $\eta = \mu_z - \lambda_z$ for the most productive blueprint, and $N_z = 0$ and $\eta > \mu_z - \lambda_z$ for all other blueprints. These possibilities are ruled out in Proposition 1 by assuming that the entrepreneurial supply of blueprints is strictly positive at all positive blueprint prices.

3.1 The Blueprint Productivity Distribution

The distribution of productivities across blueprints is determined by (18), and this is well defined because $\eta > \mu_z - \lambda_z$. Low- z blueprints may be prevalent in the economy, despite the fact that v_z/w is low, simply because many agents in the economy have the skill to introduce low- z blueprints. But high- z blueprints have the advantage that they will be replicated at a higher rate. Especially if $\mu_z - \lambda_z$ is close to η , this can easily swamp the fact that entrepreneurs introduce high- z blueprints relatively infrequently. In such a situation, the distribution of productivities can have most of its mass near the upper end of \mathcal{Z} .

3.2 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 elaborates on these two interpretations. They are benchmarks. Hybrid formulations are more plausible, but also less tractable.

3.2.1 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 (2), with $N_{z,t}$ now denoting the measure of type- z 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 labor 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 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.⁹ With this, the analysis proceeds as before.

3.2.2 Production Facilities or Plants

Instead of assuming that the output of every producer is unique, suppose there is one competitive market for final goods. Blueprints are needed to operate plants that are subject to decreasing returns, and growth in variety is no longer a source of consumption growth. Each plant can use l_t units of labor to produce output $y_t = (zQ_t l_t)^\alpha$ for some $\alpha \in (0, 1)$. The optimal scale of a type- z plant is $l_{z,t} = (\alpha z Q_t / w_t)^{\alpha/(1-\alpha)}$ units of labor. Along a balanced growth path, employment per plant is constant and the number of plants grows at the population growth rate η . It follows that wages grow at the rate

⁹There 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 ωA consumers and there are σA^2 stores that are randomly assigned to points in $[0, 1]$. The ratio of stores to consumers is σ/ω . 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.

$\kappa = \alpha\theta$, reflecting the growth rate θ of the per-capita effective labor supply, and the fact that plants are subject to decreasing returns. In contrast to the standard one-sector growth model, the cost of producing new capital rises with wages. Blueprints become more and more expensive to produce or replicate as labor becomes more efficient at operating plants. As a result, the capital stock, measured in numbers of blueprints, does not increase fast enough to allow wages to grow at the rate θ . Of course, the market value of the per-capita stock of capital does grow at the same rate as wages and output.

4. THE SIZE DISTRIBUTION OF FIRMS

The economy described up to now has agents who consume, supply labor, or act as entrepreneurs. Everyone can own blueprints and there are no firms. In this section, a transaction cost argument is used to motivate a definition of what firms are.

4.1 Transaction Costs

Consider an entrepreneur who has just developed a new blueprint. To hire labor 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 potentially very small cost 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.¹⁰

A firm will therefore only gain new commodities through its use of the technology for replicating its existing blueprints. A firm only loses commodities as its blueprints become obsolete.¹¹ A firm that has lost all its commodities is shut down. In the following, the distribution of firm size is derived assuming that the economy is on a balanced growth path.

¹⁰Of 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.

¹¹Bernard, 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.

4.2 Mechanics

For notational simplicity, assume there is only one type of blueprint, and drop the type index z . The measure of firms with n commodities at time t is denoted by $M_{n,t}$. Since every commodity is produced by one and only one firm,

$$N_t = \sum_{n=1}^{\infty} n M_{n,t}. \quad (20)$$

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

$$DM_{1,t} = \lambda 2M_{2,t} + \nu N_t - (\lambda + \mu)M_{1,t}. \quad (21)$$

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 commodity increases because firms with two commodities lose one, or because of entry. The number declines because firms with one commodity gain or lose a commodity. Similarly, the numbers of firms with more than one commodity evolve according to

$$DM_{n,t} = \lambda(n+1)M_{n+1,t} + \mu(n-1)M_{n-1,t} - (\lambda + \mu)nM_{n,t}, \quad (22)$$

for all $n-1 \in \mathbb{N}$. The joint dynamics of N_t and $\{M_{n,t}\}_{n=1}^{\infty}$ is fully described by (20)-(22).

4.3 The Stationary Size Distribution

Along the balanced growth path, N_t grows at the rate η and a stationary firm size distribution exists if (20)-(22) has 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

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

for all $n \in \mathbb{N}$. This is the fraction of firms that produce n commodities. Analytically more convenient is the fraction of all commodities produced by firms of size n , which is given by

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

for all $n \in \mathbb{N}$. Condition (20) corresponds to the requirement that the fractions Q_n add up to one,

$$1 = \sum_{n=1}^{\infty} Q_n. \quad (23)$$

Given $\nu = \eta - (\mu - \lambda)$ and the definition of $\{Q_n\}_{n=1}^\infty$, (21) can now be written as

$$\eta Q_1 = \lambda Q_2 + \eta - (\mu - \lambda) - (\lambda + \mu) Q_1, \quad (24)$$

and (22) implies that

$$\frac{1}{n} \eta Q_n = \lambda Q_{n+1} + \mu Q_{n-1} - (\lambda + \mu) Q_n, \quad (25)$$

for $n - 1 \in \mathbb{N}$. Any sequence $\{Q_n\}_{n=1}^\infty \subset [0, 1]$ that satisfies (23)-(25) defines a stationary size distribution $\{P_n\}_{n=1}^\infty$ via $Q_n \propto P_n/n$. Note that the equations (23)-(25) only depend on the parameters μ/λ and η/λ —the stationary distribution does not depend on the units in which time is measured.

PROPOSITION 2 *Suppose that μ, λ, η and $\nu = \eta - (\mu - \lambda)$ are positive. Define the sequence $\{\beta_n\}_{n=1}^\infty$ by the recursion $\beta_{n+1} = 1/(1 - (\mu\beta_n/\lambda) + (\eta + \mu n)/\lambda n)$ and the initial condition $\beta_1 = 0$. This sequence is monotone and converges to $\min\{1, \lambda/\mu\}$. The solution to (23)-(25) is given by*

$$Q_n = \frac{\nu}{\lambda} \sum_{k=n+1}^\infty \left(\prod_{m=n+1}^k \beta_m \right) \left(\frac{\mu\beta_k}{\lambda} \right)^{-1} \prod_{m=2}^k \frac{\mu\beta_m}{\lambda}. \quad (26)$$

Take any $\varepsilon > 0$. If $\mu \neq \lambda$ then

$$Q_n \in \frac{\nu}{|\mu - \lambda|} \prod_{m=2}^n \frac{\mu\beta_m}{\lambda} \times \left[\frac{1}{1 + \varepsilon}, 1 \right] \quad (27)$$

for all large enough n . The mean firm size can be written as $1/(\sum_{n=1}^\infty Q_n/n)$, and this is finite by construction.

The proof is in Appendix C. The distribution $\{P_n\}_{n=1}^\infty$ follows immediately from $P_n \propto Q_n/n$.

4.4 The Right Tail

When $\lambda > \mu$, the properties of the right-hand side of (27) are quite 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 (27) declines at a rate that is slower than any given geometric rate. The proof of Proposition 2 shows that the right-hand side of (27) is

nevertheless summable. 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 of the stationary firm size distribution satisfy*

$$\limsup_{K \rightarrow \infty} K^a \sum_{n=K}^{\infty} P_n = 0$$

for any a smaller than the tail index $\zeta = \eta/(\mu - \lambda)$.

The proof is in Appendix D. This proposition implies that

$$\ln \left(\sum_{n=K}^{\infty} P_n \right) \sim c - \zeta \ln(K) \quad (28)$$

for some constant c . The limiting tail index $\zeta = 1$ associated with Zipf's law arises when the entry rate $\nu = \eta - (\mu - \lambda)$ converges to zero.

For comparison, consider the economy of Klette and Kortum [2004]. There, $\eta = 0$ and $\mu < \lambda$. 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.

4.5 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 enters per unit of time. The firm entry rate as a fraction of the number of incumbent firms, denoted by ϕ , is therefore equal to νN_t divided by the number of firms in the economy, $\sum_{m=1}^{\infty} M_{n,t} = N_t / \sum_{m=1}^{\infty} n P_n$. An alternative way to calculate the firm entry rate ϕ is to note that the only firms that can 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 $\phi = \lambda P_1$ 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

$$\phi = \nu \sum_{n=1}^{\infty} n P_n = \eta + \lambda P_1. \quad (29)$$

Clearly, 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 (29) and $Q_1 = P_1 / \sum_{n=1}^{\infty} n P_n$ imply $\phi/\eta = 1/(1 - (\lambda/\nu)Q_1)$. Together with (26) 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 mean firm size ϕ/ν .

4.6 A Convenient Limiting Case

Suppose η and λ are bounded away from zero and let $\mu - \lambda$ approach η from below so that the rate ν at which blueprints are introduced by entrepreneurs goes to zero. Observe that this is exactly when the tail index $\zeta = \eta/(\mu - \lambda)$ approaches one from above. In the limit, the recursion (25) for $Q_n \propto n P_n$ can be written as $P_n = \frac{\lambda}{\mu} (P_{n+1} + X_{n+1})$ together with $X_{n+1} = \binom{n-1}{n+1} X_n$ for all $n - 1 \in \mathbb{N}$. This implies $X_{n+1} = (2/[n(n+1)])X_2$ for all $n \in \mathbb{N}$. Iterating forward on the recursion for P_n and requiring the resulting P_n to add up to one yields

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

This distribution does not have a finite mean. The implied right tail probabilities satisfy

$$\lim_{K \rightarrow \infty} K \sum_{n=K}^{\infty} P_n = \lim_{K \rightarrow \infty} \frac{1}{\ln(\mu/\eta)} \sum_{m=0}^{\infty} \frac{K}{K+m} \left(\frac{\lambda}{\mu}\right)^{m+1} = \frac{1}{\ln(\mu/\eta)} \frac{1}{\mu/\lambda - 1}$$

by the dominated convergence theorem (see Appendix B.) Thus the right tail probabilities behave like $1/K$, and the log right tail probabilities expressed as a function of $\ln(K)$ must asymptote to a straight line with slope -1 .

At the same time as the rate ν at which blueprints are introduced by entrepreneurs goes to zero, the average number of blueprints per firm goes to infinity. But the entry rate of new firms satisfies $\phi = \eta + \lambda P_1$, and this converges to a positive value. A calculation yields $\phi = \lambda/\ln(\mu/\eta)$, and then $\mu - \lambda = \eta$ gives

$$\frac{\phi}{\eta} = \frac{\mu/\eta - 1}{\ln(\mu/\eta)}. \quad (30)$$

In situations when most blueprints are created by incumbent firms, this allows one to infer μ/η and $\lambda/\eta = -1 + \mu/\eta$ simply from the ratio of the firm entry rate over the population growth rate ϕ/η .

4.7 Heterogeneous Blueprint Qualities and Gibrat's Law

A firm is defined as the collection of blueprints produced by replication from some initial blueprint. Replication produces blueprints of the same quality and so one can take the type of a firm to be the uniform quality of its blueprints. Propositions 2 and 3 then apply to firms of the same type.

Along the balanced growth path, firm growth rates satisfy Gibrat's law conditional on type, and the size distribution of type- z firms will have a tail index $\eta/(\mu_z - \lambda_z)$. Since high-type blueprints are replicated more quickly than low-type blueprints, this implies a thicker right tail for the distribution of high-type firms. The right tail of the overall size distribution will inherit the tail index of the highest firm type, which corresponds to the lowest $\eta/(\mu_z - \lambda_z)$. Most firms far out in the right tail of the distribution will be high- z firms. Since these are the firms with the highest mean growth rate in the economy, this induces a positive association between firm size and firm growth rates.

Most empirical studies suggest that Gibrat's law is violated in the other direction: small firms tend to have higher growth rates than medium or large firms; see Evans [1987], Hall [1987], Dunne, Roberts and Samuelson [1989], and the survey of Sutton [1997]. At the root of the problem is the assumption that firms have one type of blueprint and that this type is permanent. As a consequence, mean growth rates are permanently different across firms. The data shown in Figure I suggest, instead, that firms initially grow at rates that far exceed the restriction $\mu_z - \lambda_z < \eta$ implied by Proposition 1, and that these growth rates decline as firms age.

4.8 Firm Type Transitions

A simple way to account for this slow-down in firm growth rates and examine its implications for the stationary size distribution is as follows. Suppose that $\mathcal{Z} = \{z_L, z_H\}$ for some $z_L < z_H$, and that new firms enter as type- z_H firms and then transition to type- z_L firms following independent and exponentially distributed waiting times with mean $1/\delta$. When such transitions happen, all blueprints of a firm turn into type- z_L blueprints, permanently. Any new blueprints created by the firm thereafter will be of type z_L .

The incentives to invest continue to be determined by (17) along a balanced growth path. But the present-value condition for blueprint values (16) in the initial phase must

be modified to account for the loss in value that occurs when a blueprint transitions from high quality to low quality. This yields

$$\frac{v_{z_H}}{w} = \frac{\psi[l_{z_H} - [i_{z_H} + j_{z_H}] - \delta \frac{[v_{z_H} - \frac{v_{z_L}}{w}]}{w}]}{r - \kappa - [\mu_{z_H} - \lambda_{z_H}]}.$$

In the terminal phase, v_{z_L}/w continues to satisfy (16). Along the balanced growth path, the number of type- z_H blueprints is $N_{z_H}/H = E_{z_H}(v, w)/(\eta + \delta - [\mu_{z_H} - \lambda_{z_H}])$. Type- z_L blueprints are created by replication and because a flow δN_{z_H} of type- z_H blueprints depreciate in quality. The resulting number of type- z_L blueprints is therefore $N_{z_L}/H = \delta(N_{z_H}/H)/(\eta - [\mu_{z_L} - \lambda_{z_L}])$. Given this, the labor market clearing condition (19) becomes

$$\frac{E_{z_H}(v, w)}{\eta + \delta - [\mu_{z_H} - \lambda_{z_H}]} \left(l_{z_H} + i_{z_H} + \frac{\delta[l_{z_L} + i_{z_L}]}{\eta - [\mu_{z_L} - \lambda_{z_L}]} \right) = L(v, w).$$

If the talent distribution for entrepreneurs is unbounded, $E_{z_H}(v, w)$ is positive at all positive prices. This then implies that $\eta + \delta > \mu_{z_H} - \lambda_{z_H}$ and $\eta > \mu_{z_L} - \lambda_{z_L}$ in equilibrium. Using the first of these inequalities together with $r - \kappa > \eta$ one can apply the same argument as before to show that optimality implies that $i_{z_H} > i_{z_L}$ and $j_{z_H} > j_{z_L}$. While their quality advantage lasts, type- z_H firms have stronger incentives to invest in replicating blueprints than type- z_L firms. They lose blueprints at a lower rate and gain new ones at a higher rate. Firms grow faster in the initial phase.

PROPOSITION 4 *Suppose firms enter with productivity z_H and transition to productivity $z_L < z_H$ at a positive rate δ . Then, along the balanced growth path, $\mu_{z_H} \geq \mu_{z_L}$, $\lambda_{z_H} < \lambda_{z_L}$, $\eta + \delta > \mu_{z_H} - \lambda_{z_H}$ and $\eta > \mu_{z_L} - \lambda_{z_L}$. The stationary size distribution has a tail index ζ given by*

$$\zeta = \frac{\eta + \delta}{\mu_{z_H} - \lambda_{z_H}} \quad \text{if} \quad 1 < \frac{\eta + \delta}{\mu_{z_H} - \lambda_{z_H}} < \frac{\eta}{[\mu_{z_L} - \lambda_{z_L}]^+}$$

and

$$\zeta = \frac{\eta}{\mu_{z_L} - \lambda_{z_L}} \quad \text{if} \quad 1 < \frac{\eta}{\mu_{z_L} - \lambda_{z_L}} \leq \frac{\eta + \delta}{\mu_{z_H} - \lambda_{z_H}}.$$

The right tail of the size distribution declines geometrically otherwise.

The actual size distribution and a proof of Proposition 4 are in Appendix E. If $\zeta = \eta/(\mu_{z_L} - \lambda_{z_L})$ then the fact that all firms eventually grow at a rate that is close to the population growth rate accounts for the thick right tail observed in the firm size distribution. Alternatively, if $\zeta = (\eta + \delta)/(\mu_{z_H} - \lambda_{z_H})$, 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. In an extreme example, one can set δ slightly above $\mu_{zH} - \lambda_{zH}$ and make both arbitrarily large. New firms then either exit or become large almost instantaneously.

One possible interpretation for the firm type transitions discussed here is that some aspect of the environment for which the initial blueprint of a firm was created has changed permanently.¹² Of course, such changes could take place more gradually than described here. 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.

5. FIRM AGE AND SIZE

The early histories of large firms are often known in some detail. This makes the age distribution among large firms a useful tool for assessing alternative interpretations of the firm size distribution. This section derives this distribution and the next will use it to show that Gibrat’s law is inconsistent with the data.

5.1 The Size Distribution of a Cohort

Consider a cohort of firms that enter at the same time, not necessarily with a single blueprint. As in Section 4, suppose these firms are initially in a “phase one” in which they gain and lose blueprints at certain rates. Firms that have not exited transition into a “phase two” following an exponentially distributed waiting time with mean $1/\delta$. Let $s_{-1}(a)$ denote the fraction of firms in the cohort that have made this transition by age a and define $s_n(a)$ to be fraction of all cohort firms that are in phase one and have n blueprints. In phase one, firms with n blueprints gain new blueprints at a rate μn and lose existing blueprints at a rate λn . In particular, firms that lose their last blueprint will remain zero-blueprint firms forever—they exit. Since only firms in phase one that have not yet exited can transition into phase two,

$$Ds_{-1}(a) = \delta [1 - s_{-1}(a) - s_0(a)]. \quad (31)$$

¹²Atkeson and Kehoe [2006] interpret the observed slow-down in firm growth rates by assuming that firm productivity growth rates decline with age. Here the interpretation is a decline in the level of productivity relative to everyone else. A plausible cause for such a decline is competing firms catching up with a firm’s initial innovation. Although the economics is quite clear, describing this in a way that yields an analytically tractable growth process is more challenging.

Exits occur when firms lose their last blueprint, and hence

$$Ds_0(a) = \lambda s_1(a). \quad (32)$$

The number of firms of a cohort that are still in phase one and that have n blueprints by age a must satisfy

$$Ds_n(a) = (n-1)\mu s_{n-1}(a) + (n+1)\lambda s_{n+1}(a) - [\delta + (\mu + \lambda)n]s_n(a) \quad (33)$$

for all $n \in \mathbb{N}$. Note that the $-\delta s_n(a)$ term is not scaled by n , reflecting the assumption that the transition probability from phase one to phase two is independent of size. The probability distribution $\{s_n(a)\}_{n=-1}^{\infty}$ is determined by (31)-(33) and an initial condition for the size distribution of a cohort at entry.

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 $k \in \mathbb{N}$ and for any $\delta \geq 0$ define*

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

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 $n \in \mathbb{N}$. Then $s_n(a) = \mathcal{T}_{n,k}(a)$ solves (31)-(33) for the initial condition given by $s_1(0) = k$ and $s_n(0) = 0$ otherwise.

For $\delta = 0$ and $k = 1$ this solution can be found and 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 (32). 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 (31). The proof of Proposition 5 can be completed by computing the derivative of $\mathcal{T}_{n,k}(a)$ and checking (33) for any $n \in \mathbb{N}$. Appendix F gives a more constructive proof based on the observation that, conditional on no transition from phase one to phase two, 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.

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

5.2 Age Given Size

Now consider the setup of Proposition 4: a cohort of firms initially produces goods with a blueprint quality z_H , and that this quality depreciates permanently to z_L following an exponentially distributed waiting time with mean $1/\delta$. Write $\mathcal{T}_{H,n,k}(a)$ and $\mathcal{T}_{L,n,k}(a)$ for the solutions to (31)-(33) associated with $(\mu_{z_H}, \lambda_{z_H}, \delta)$ and $(\mu_{z_L}, \lambda_{z_L}, 0)$, respectively. Then the cohort size distribution $\{p_n(a)\}_{n=0}^{\infty}$ at age a is given by

$$p_n(a) = \mathcal{T}_{H,k,1}(a) + \delta \int_0^a \left[\sum_{k=1}^{\infty} \mathcal{T}_{L,n,k}(b) \mathcal{T}_{H,k,1}(a-b) \right] db \quad (34)$$

for all $n + 1 \in \mathbb{N}$. The infinite sum on the right-hand side of (34) can be calculated explicitly, as reported in Appendix F. The first term on the right-hand side of (34) accounts for the firms that are still in the initial growth phase. The second term represents firms that have transitioned into the second phase by age a . A flow $\delta \mathcal{T}_{H,k,1}(a-b)$ of firms in the initial growth phase transition into the second phase at age $a-b$. Adding up over all ages and accounting for their subsequent growth gives the second term. Note well that (34) includes $n = 0$, and so the sum of $p_n(a)$ over all $n \in \mathbb{N}$ gives the fraction of firms that have survived up to age a .

Along a balanced growth path, the measure of new firms entering is growing at a rate η . Consider the population of all firms that have entered up to a particular point in time, including those that have since exited. The exponential rate η at which entry cohort sizes grow implies an exponential age distribution $1 - e^{-\eta a}$ for this population. Because $\{p_n(a)\}_{n=0}^{\infty}$ 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 $n \geq K$ is then

$$h_K(a) = \frac{\sum_{n=K}^{\infty} \eta e^{-\eta a} p_n(a)}{\sum_{n=K}^{\infty} \int_0^{\infty} \eta e^{-\eta b} p_n(b) db}. \quad (35)$$

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

6. 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 (Arrington [1998, p.9]). This is likely to create significant biases in measured employment for small firms. The SBA reports firm counts for 24 size categories, ranging from 1 to 4 employees to 10,000 and more employees, as well as the number of employer firms with no employment in March but some employment at other times during the year. Over the 1990's, SBA data show that the number of firms grows roughly at the population growth rate of about 1% per annum, as the model predicts.

Age data on firms with more than 10,000 employees in 2006 will also be used. Two measures of firm age are reported. 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.

6.1 Gibrat Implies 750 Year Old Firms

Panels (i) and (ii) of Figure III show the fitted employment size distribution assuming there is only one growth phase. The fractions $\#\{\text{firms with employment} \leq n\}/\#\{\text{all firms}\}$ and $\#\{\text{firms with employment} \geq n\}/\#\{\text{all firms}\}$ 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 $n^{-\zeta}$, and the slope of the log tail probabilities with respect to n is about $\zeta \approx 1.05$. Note that this estimate does not depend on the units in which firm employment is measured. Annual U.S. population growth is around 1%. The formula for the tail index ζ then implies that firms grow at a rate $\mu - \lambda = \eta/\zeta \approx .95\%$ per annum.

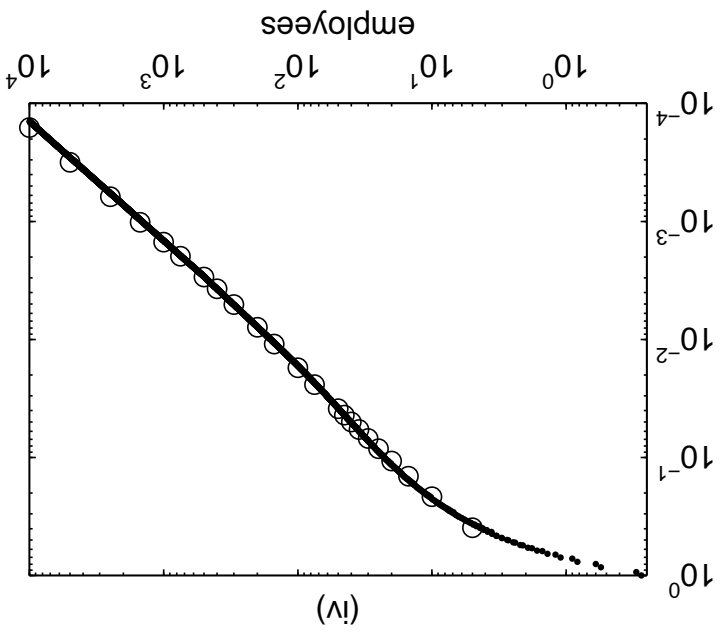
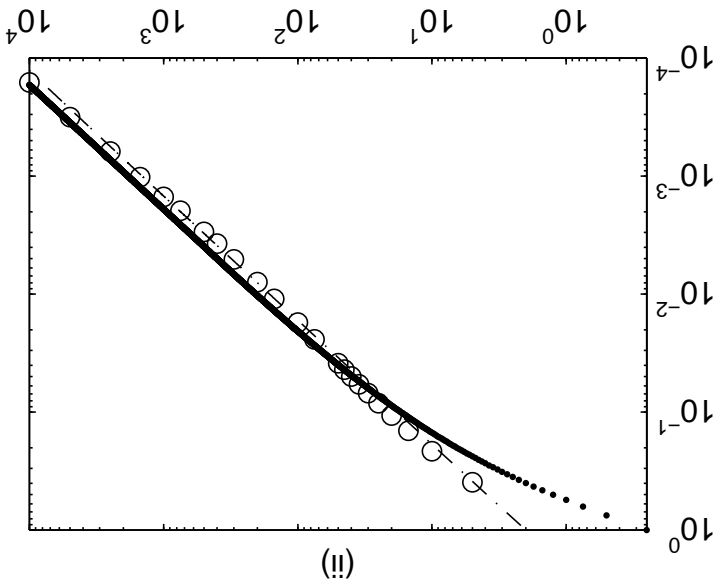
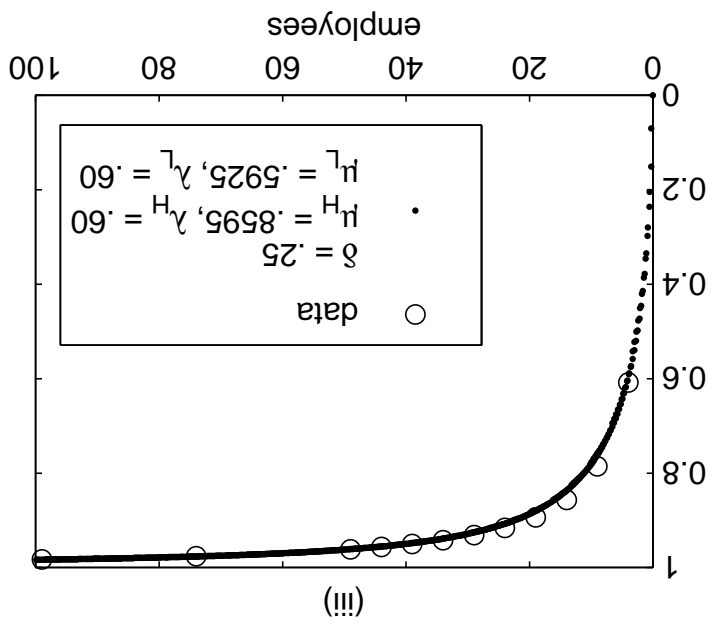
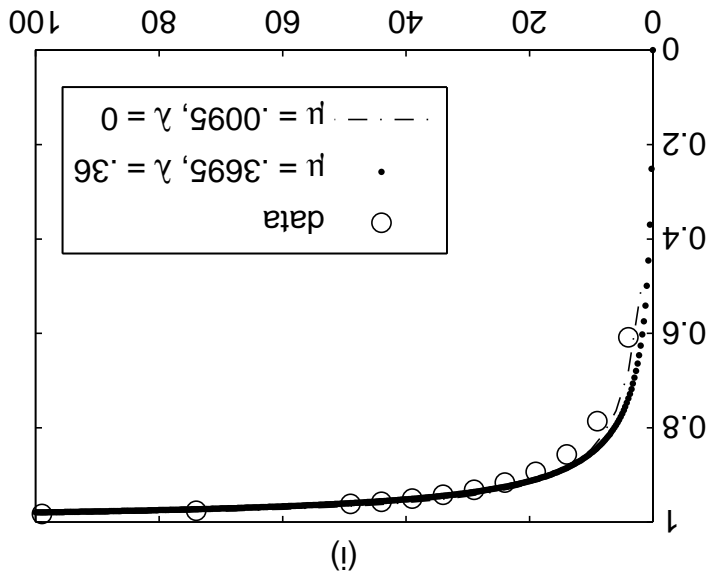
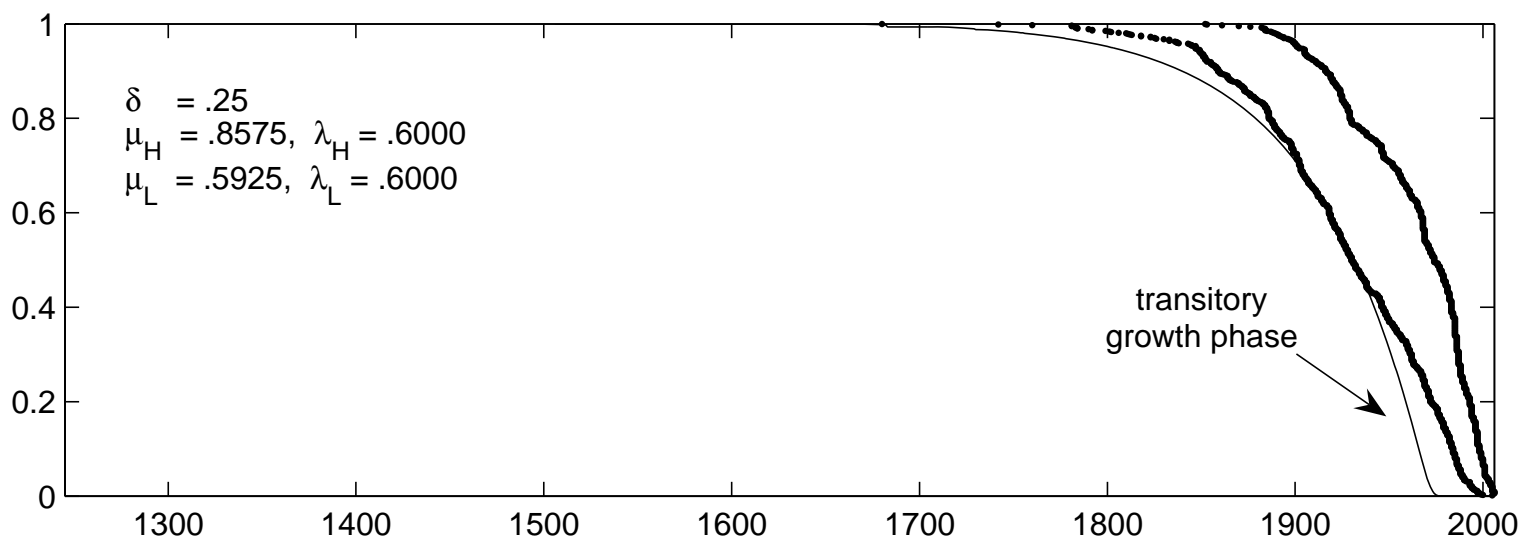
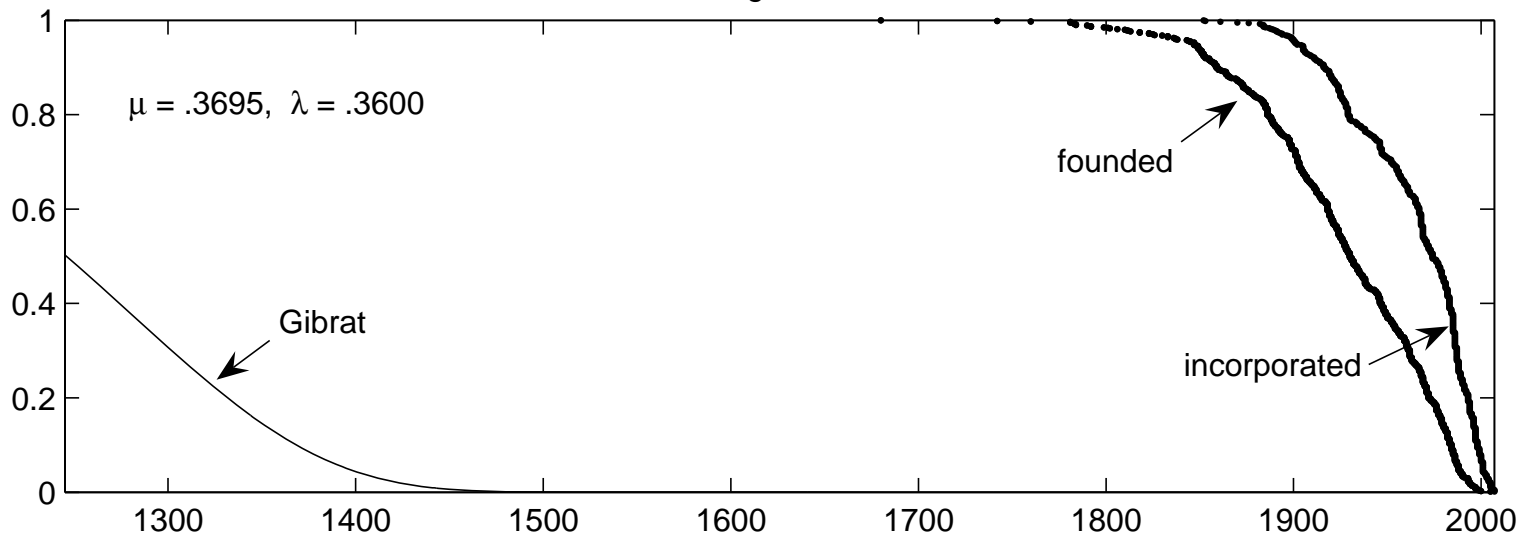


Figure III

Figure IV



To decompose $\mu - \lambda$, consider first the Yule process obtained by setting $\lambda = 0$ and $\mu = .0095$. The only remaining free parameter is then the number of employees per blueprint $i + j + l$. Changing this parameter causes parallel shifts in the curves representing the model shown in panel (ii) of Figure III. 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 this model of firm growth has deficiencies that show up very clearly in dynamic data. The Yule process predicts a firm entry rate $\phi = \eta + \lambda P_1 = \eta$, and this equals only about 1% per annum. Instead, the SBA reports a firm entry rate of about 10% per annum over the 1990's. Actual firms do decline and exit, and entry rates are much higher than the population growth rate.

To match the evidence on firm entry along with the shape of the right tail of the size distribution, one can raise μ and λ subject to the constraint $\mu - \lambda = \eta/\zeta \approx .0095$ until the implied entry rate ϕ reaches the .1 value observed in the data. Solving the $\zeta \downarrow 1$ approximation (30) of the firm entry rate gives $\mu = .3615$ and $\lambda = .3615$. A precise calculation for $\zeta = 1.05$ gives $\mu = .3695$ and $\lambda = .36$. Choosing the number of employees per blueprint to match the right tail of the size distribution now gives $i + j + l = .25$. The associated left and right tails are again shown in panels (i) and (ii) of Figure III. 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 IV show a much more dramatic failure of the one-phase model of firm growth. At $\mu = .3695$, $\lambda = .36$ and $i + j + l = .25$, 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 closer to 75 years. Given Gilbrat's law, firms all grow at the same average rate, and this must be equal to $\mu - \lambda = \eta/\zeta \approx .0095$. At this pace, it takes a very long time for a firm to grow from its initial size of $i + j + l = .25$ employees to one with 10,000 employees—perhaps not quite the $\ln(40,000)/.0095 \approx 1,115$ years implied by deterministic growth, but still, a very long time.¹³

¹³The back-of-the-envelope age of 1,115 years can be reduced to 485 years by assuming that firms

6.2 Rapid Initial Growth

As Figure I suggests, many large firms started out small and became large during relatively short periods of rapid growth, at rates far exceeding the population growth rate. 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 to a regime with a growth rate $\mu_L - \lambda_L$ that must be below η . If the tail index is determined by the effects of initial rapid growth, then $\zeta = (\eta + \delta)/(\mu_H - \lambda_H)$. Given $\zeta \approx 1.05$ and $\eta \approx .01$, this implies that $\mu_H - \lambda_H$ must be close to δ . An initial phase with very rapid growth is possible as long as this phase is of sufficiently short average duration. The fact that transition times are exponentially distributed implies that some firms grow rapidly for much longer than the average duration. This results in relatively young large firms.

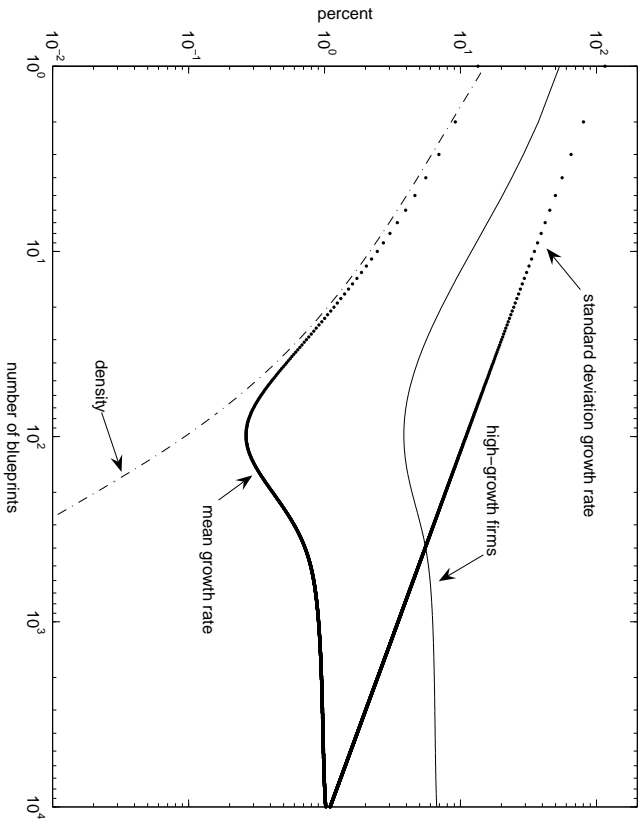


FIGURE V Growth by Size

enter with 100 employees and grow deterministically. This still implies large firms that are much too old. In Luttmer [2007], Gibrat's law holds in a strong form: both the mean and the variance of firm growth rates are independent of size. The fact that the variance does not decrease with size makes it easier for selection to produce large firms. But it takes annual growth rates with a standard deviation as high as 40% per annum, implausibly high for all but the smallest firms, to account for the observed age distribution of large firms.

Panels (iii) and (iv) of Figure III and the lower panel of Figure IV show the size and age distributions when $\mu_H = .8575$, $\mu_L = .5925$, and $\lambda_H = \lambda_L = .6$, together with $\delta = .25$ and employee numbers per blueprint equal to $i_H + j_H + l_H = .30$ and $i_L + j_L + l_L = .275$. These parameters give rise to a firm entry rate of 10% and a median age among firms with more than 10,000 employees of 75 years. The parameters were chosen to match these two features of the data, and to approximate the left and right tails of the size distribution, subject to the theoretical restrictions implied by the model. Roughly, the mean growth rates $\mu_H - \lambda_H$ and $\mu_L - \lambda_L$ are important for the median age, and the variance parameters $\mu_H + \lambda_H$ and $\mu_L + \lambda_L$ are important for the entry rate. Increasing $\mu_H - \lambda_H$ lowers the median age while increasing $\mu_L - \lambda_L$ causes the median age to rise as firms that have become large stay around longer. Increases in $\mu_H + \lambda_H$ and $\mu_L + \lambda_L$ raise the entry rate as firms spread out and exit more quickly. If $i_H + j_H + l_H$ and $i_L + j_L + l_L$ are reduced along with these increases to keep the frequency of large firms in line with the data, then the median age of large firms does not change much.

The two-phase model introduces variation in mean growth rates that is correlated with firm size. Figure V shows the mean and standard deviation of firm growth conditional on size, with size measured by the number of blueprints employed by the firm. A logarithmic scale is used to accentuate variation over the full range of the data. The graph shows that beyond 300 blueprints, or about 100 employees, the implied mean firm growth rate is essentially constant. This can explain why many researchers have found that Gibrat's law is a good approximation for firms that are not too small (e.g. Hall [1987] and Evans [1987].) Small firms grow much faster on average, and the model also predicts an intermediate size range in which firms are most likely to be in the low-growth phase.

Over short intervals of time, the variance of firm growth in regime i is $(\mu_i + \lambda_i)/n$ for a firm with n blueprints. The resulting estimated standard deviation of a firm with n blueprints is about $1.21/\sqrt{n}$ in the high-growth phase and $1.09/\sqrt{n}$ in the low-growth phase. A firm with 30 employees has around 100 blueprints, and hence the standard deviation of its growth rate will be in the 10% – 12% range. For firms with more than 10,000 employees, these standard deviations are on the order of .5% to .6%, numbers that are small even relative to aggregate employment fluctuations in the U.S. economy. 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 Computstat firms behaves like $1/n^{1/3}$. Tentative interpretations are given in Stanley et al. [1996] and Sutton [2002, 2007].

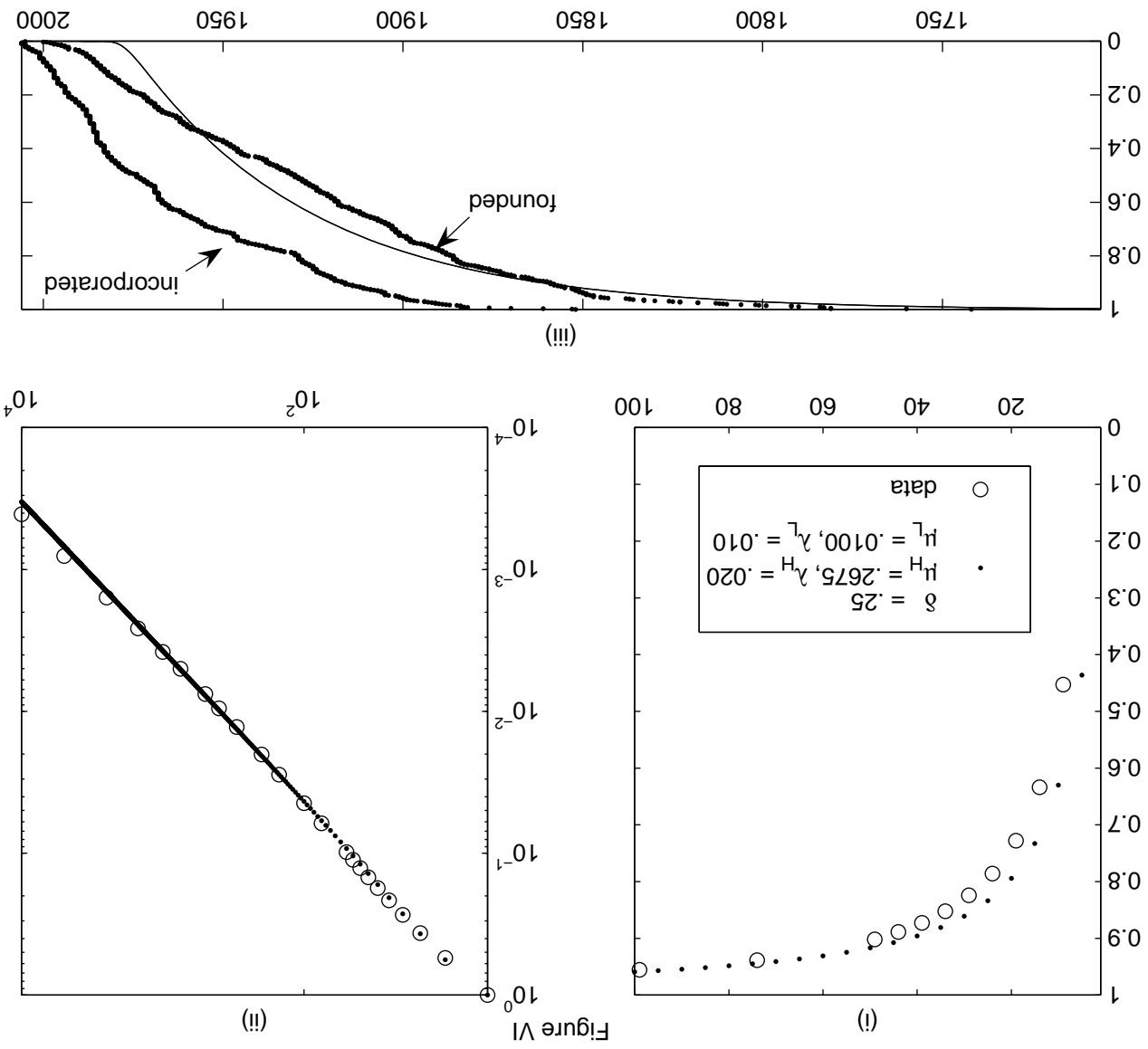
6.3 What Could Those Blueprints Be?

For the Yule process, $i + j + l = 2$ is the level of employment per blueprint that best fits the empirical employment size distribution. If μ and λ are increased to match the observed firm entry rate, then employment per blueprint drops to $i + j + l = .25$. Similar estimates of employment per blueprint are obtained if there is an initial growth phase. Increasing employment per blueprint and increasing the variance of the firm growth process both cause the stationary employment distribution to spread out. The observed dispersion of employment therefore puts tight restrictions on employment per blueprint and the variance of the firm growth process. If the employment per blueprint is high, then the variance of the growth rate must be low, and vice versa. The Yule process has the lowest variance consistent with the observed tail index, and hence the highest level of employment per blueprint.

Therefore, if the model is to fit the distribution of all firms in the SBA collection of firms, employment per blueprint can be at most 2. Such a low number rules out several a priori plausible interpretations of what blueprints could be. Plants, organizational departments, or geographically distinct subsidiaries typically have more than two employees. Projects, tasks, sales accounts, or individual job descriptions might be better interpretations. Strictly within the model, blueprints are associated with differentiated commodities, and the small number of employees assigned to each differentiated commodity suggests an extremely differentiated set of commodities.

Figure VI shows an alternative way to interpret part of the SBA data. The empirical distribution displayed in this figure is the distribution of all firms with at least five employees. That is, firms with 0 employees in March and those with 1-4 employees are not included. Within the context of the model, one could interpret these small employer firms as intermediaries that supply labor services to organizations that are classified as firms—those that have more than five employees. Conditioning on firms with at least five employees causes the log of the empirical right tail frequencies reported in panel (ii) to shift up by a constant. This is what allows one to match the data with a larger number of employees per blueprint. In Figure VI, this is taken to be the minimum size of 5 employees. As before, $\delta = .25 \approx \mu_H - \lambda_H$, but $\mu_H + \lambda_H$ and $\mu_L + \lambda_L$ now have to be smaller to match the right tail of the distribution. Panel (iii) shows that the

median large firm is now younger than the observed 75 years, while other percentiles are above what is found in the data. The overall age distribution is still in the range of the empirical age distribution.



7. CONCLUSION

Data on U.S. firms show (i) a size distribution with many small firms and a very thick right tail, (ii) high firm entry and exit rates, (iii) relatively young large firms, and (iv) growth rate variances that decline with size more slowly than the inverse of firm size. Luttmer [2007] and the current paper provide alternative interpretations of (i). Large amounts of randomness or high transitory growth rates are required to account for (ii) and (iii). The two models seem to be on opposite sides of the data when it comes to (iv). In one case all shocks are firm-wide, while in the other independent within-firm shocks average out. But neither model has industry or aggregate shocks. Evidence on the slow decline of the variance of firm growth rates tends to come from sources, such as Compustat, that do not include very small firms. It is possible that this misses a rapid initial decline with size in the growth rate variance of these firms.

Skewed firm size distributions are interpreted as reflecting skewed productivity distributions in Hopenhayn [1992], Atkeson and Kehoe [2006], 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 [2007], and Arkolakis [2007]. 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 Terzio [2007], both firm-specific productivity and managerial productivity play a role. Much remains to be done to sort out the relative importance of each 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 all 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. Of course, not all new firms have high-quality blueprints. The framework set out in this paper allows for heterogeneous quality in the blueprints of start-up firms, and this can account for small firms that linger and never “take off.” 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 I, 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 I would simply reflect the fact that setting up a large firm is not quite instantaneous but still very fast.

A close examination of the early histories of large U.S. corporations, such as those shown in Figure I 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 [2004], 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 how these aspects of firm growth can be incorporated in a model of the size distribution. And of course, these phenomena can shed light on what keeps the components of a firm together.

A FIRM AGE DATA

The firm age data for large firms used in Section 7 is collected from several sources. Large firms are taken to be all Compustat firms with more than 10,000 employees, about 600 companies, together with about 300 firms in that same size category that appear on a list of large private companies published by *Forbes* magazine. For these firms, two age measures were collected. One is the incorporation date obtained from the Merger data base. The second measure is the earliest reference to the company, or any of its predecessor companies, that can be found in any of three different sources: Merger (formerly Moody's manual), Dun and Bradstreet, and company web sites. A predecessor company can be a company that was broken up into parts, or it can be a component of the company that was initially independent. For firms that are in the Merger database, fairly extensive histories are reported, and these histories contain information about predecessor companies. Company web sites of large corporations often include extensive company histories that tend to emphasize the very old roots of the company. The company age data together with the source for each age observation are available at www.luttmer.org.

B THE LIMITING CASE $\zeta \downarrow 1$

Iteration gives

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

for some ξ . Note that

$$\begin{aligned} \frac{1}{\xi} \sum_{n=K}^{\infty} P_n &= \sum_{n=K}^{\infty} \sum_{k=n}^{\infty} \frac{(\lambda/\mu)^{k+1-n}}{k(k+1)} \\ &= \sum_{m=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^{m+1} \sum_{n=K}^{\infty} \frac{1}{(m+n)(m+n+1)} = \sum_{m=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^{m+1} \frac{1}{m+K}. \end{aligned}$$

For $K = 1$ this gives

$$\frac{1}{\xi} \sum_{n=1}^{\infty} P_n = \sum_{m=1}^{\infty} \frac{1}{m} \left(\frac{\lambda}{\mu}\right)^m = -\ln\left(1 - \frac{\lambda}{\mu}\right).$$

Since $\eta = \mu - \lambda$ it follows that $\xi = -1/\ln(1 - \lambda/\mu) = 1/\ln(\mu/\eta)$. The firm entry rate is then

$$\phi = \eta + \lambda P_1 = \eta + \frac{\lambda}{-\ln\left(1 - \frac{\lambda}{\mu}\right)} \sum_{k=1}^{\infty} \frac{(\lambda/\mu)^k}{k(k+1)} = \lambda \xi,$$

which is the result reported in (30).

C PROOF OF PROPOSITION 2

Equation (25) is a second-order difference equation in $\{Q_n\}_{n=1}^{\infty}$. It comes with two boundary conditions, (23) and (24). To solve this system, it is convenient to introduce a sequence $\{\beta_n\}_{n=1}^{\infty}$ that reduces (25) to a first-order equation in the variables

$$Z_{n+1} = \frac{1}{\beta_{n+1}} [Q_n - \beta_{n+1} Q_{n+1}], \quad (36)$$

for all $n \in \mathbb{N}$. Specifically, set $\beta_1 = 0$ and

$$\beta_{n+1} = \left(1 + \frac{\eta + \mu n}{\lambda n} - \frac{\mu \beta_n}{\lambda}\right)^{-1} \quad (37)$$

for all $n \in \mathbb{N}$. Then the boundary condition (24) translates into

$$Z_2 = \frac{1}{\lambda} [\eta - (\mu - \lambda)] \quad (38)$$

and (25) can be written as

$$Z_{n+1} = \left(\frac{\mu\beta_n}{\lambda} \right) Z_n \quad (39)$$

for all $n - 1 \in \mathbb{N}$. The recursion (37) is depicted in Figure II for the case $\mu > \lambda$.

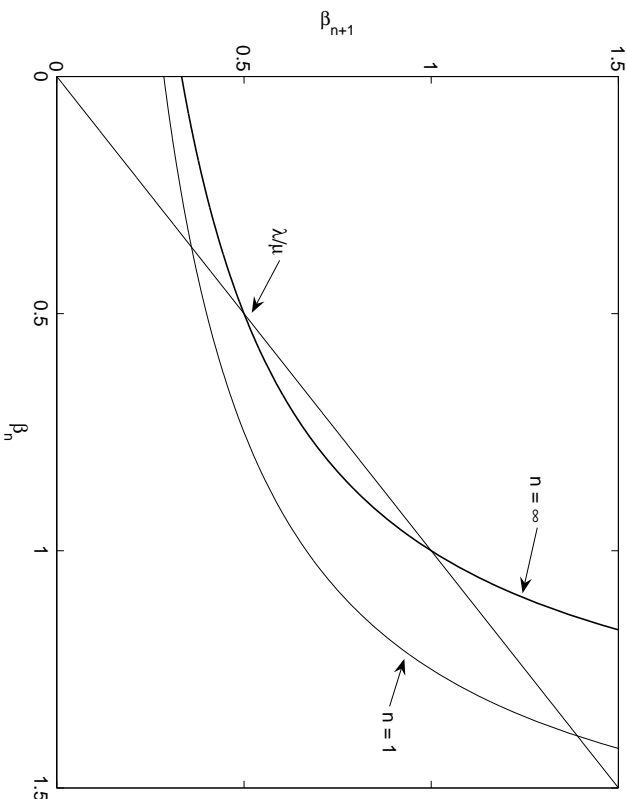


FIGURE II. The Dynamics of $\{\beta_n\}_{n=1}^\infty$.

Note in particular that the curve defined by (37) shifts upwards as n increases. Using this observation one can verify that $\{\beta_n\}_{n=1}^\infty$ converges monotonically from $\beta_1 = 0$ to $\min\{1, \lambda/\mu\}$. The sequence $\{Z_n\}_{n=2}^\infty$ is completely determined by (38)-(39). Observe from (36) that $Q_n = \beta_{n+1}(Q_{n+1} + Z_{n+1})$. The boundary condition (23) together with the fact that $\beta_n \leq 1$ implies that $Q_K \prod_{k=1}^K \beta_k$ must converge to zero as K becomes large. Thus one can iterate forward to obtain the solution for $\{Q_n\}_{n=1}^\infty$. The solution is given in the statement of Proposition 1.

Write the candidate solution (26) as

$$\left(\prod_{m=2}^n \frac{\mu\beta_m}{\lambda} \right)^{-1} Q_n = \nu \frac{1}{\lambda} \sum_{k=n+1}^\infty \left(\prod_{m=n+1}^k \beta_m \right) \left(\frac{\mu\beta_k}{\lambda} \right)^{-1} \prod_{m=n+1}^k \frac{\mu\beta_m}{\lambda}$$

Since $\beta_n \leq \min\{1, \lambda/\mu\}$ this implies the upper bound in (27). Take some $\varepsilon > 0$. The lower bounds rely on $\beta_n \uparrow \min\{1, \lambda/\mu\}$. If $\mu > \lambda$, then eventually $\beta_n \geq (\lambda/\mu)/(1 + \varepsilon)$, and this gives the lower bound in (27). If $\mu < \lambda$, then $\beta_n \geq 1/(1 + \varepsilon)$ for all large enough n , and this implies (27). Thus the sums defining $\{Q_n\}_{n=1}^\infty$ converge and (27)

holds. By construction, the candidate solution satisfies (24)-(25). It remains to prove the adding-up condition (23).

Define $F_1 = 1$ and

$$F_n = n \prod_{k=2}^n \frac{\mu \beta_k}{\lambda},$$

for all $n - 1 \in \mathbb{N}$. Note from the bound (27) that the sequence $\{Q_n\}_{n=1}^\infty$ is summable if and only if $\{F_n/n\}_{n=1}^\infty$ is summable. This summability and $F_n \rightarrow F_\infty = 0$ are immediate for $\mu < \lambda$. Define

$$X_n = n \left(1 - \frac{\mu \beta_n}{\lambda} \right)$$

for all $n \in \mathbb{N}$. The recursion (37) for β_n is equivalent to

$$X_{n+1} = \frac{\left(1 + \frac{1}{n}\right) \left(X_n + \frac{\eta}{\lambda}\right)}{\frac{\mu}{\lambda} + \frac{1}{n} \left(X_n + \frac{\eta}{\lambda}\right)}. \quad (40)$$

Starting from $X_1 = 1$, this converges monotonically to $\eta/(\mu - \lambda) > 1$ if $\mu > \lambda$. Raabe's test then ensures that $\{F_n/n\}_{n=1}^\infty$ is summable. The inequality $X_{n+1} > 1$ is equivalent to $F_{n+1} < F_n$, and so $F_n \downarrow F_\infty$ for some $F_\infty \geq 0$. Hence

$$\sum_{k=1}^K \frac{1}{k} F_k \geq F_\infty \sum_{k=1}^K \frac{1}{k}$$

for all K . Since the left-hand side is summable, it must be that $F_\infty = 0$.

Write (24) as $\eta Q_1 = \lambda [Q_2 - Q_1] + \eta - (\mu - \lambda) - \mu Q_1$ and (25) as

$$\eta Q_n = \lambda n [Q_{n+1} - Q_n] - \mu n [Q_n - Q_{n-1}],$$

for $n - 1 \in \mathbb{N}$. Adding up gives over all n gives

$$\eta \sum_{n=1}^{\infty} Q_n = \eta - (\mu - \lambda) + \lambda \sum_{n=1}^{\infty} n [Q_{n+1} - Q_n] - \mu \left(Q_1 + \sum_{n=2}^{\infty} n [Q_n - Q_{n-1}] \right). \quad (41)$$

Note that $n[Q_{n+1} - Q_n] = (n + 1)Q_{n+1} - nQ_n - Q_{n+1}$ and $n[Q_n - Q_{n-1}] = nQ_n - (n - 1)Q_{n-1} - Q_{n-1}$, and observe that $F_\infty = 0$ ensures that the candidate solution (26) satisfies $\lim_{n \rightarrow \infty} nQ_n = 0$. Using summation-by-parts for the two sums on the right-hand side of (41) one obtains

$$\sum_{n=1}^{\infty} n [Q_{n+1} - Q_n] = Q_1 + \sum_{n=2}^{\infty} n [Q_n - Q_{n-1}] = - \sum_{n=1}^{\infty} Q_n. \quad (42)$$

Together with $\eta > \mu - \lambda$, (41) and (42) imply that the sequence $\{Q_n\}_{n=1}^\infty$ adds up to 1.

The infinite sum (26) is hard to compute accurately when ζ is close to 1. The computations in Section 6 make use of the recursion $Q_n = \beta_{n+1}(Q_{n+1} + Z_{n+1})$ and the bound (27) for large n .

D PROOF OF PROPOSITION 3

Recall that $P_n \sim F_n/n^2$ and define $R_K = \sum_{k=K}^{\infty} F_k/k^2$. Observe

$$K^a R_K = \sum_{k=K}^{\infty} \left(\frac{K}{k}\right)^a k^{a-1} \prod_{m=2}^k \frac{\mu \beta_m}{\lambda} \leq \sum_{k=K}^{\infty} k^{a-1} \prod_{m=2}^k \frac{\mu \beta_m}{\lambda}.$$

If the sum on the right-hand side is finite, then $\lim_{K \rightarrow \infty} K^a R_K = 0$. A sufficient condition for this sum to converge is a version of Raabe's test, $\lim_{n \rightarrow \infty} Y_n > 1$, where

$$Y_n = n \left(\left[\frac{n-1}{n} \right]^{a-1} \frac{\lambda}{\mu \beta_n} - 1 \right).$$

The recursion (37) for β_n is equivalent to

$$Y_{n+1} - A_{n+1} = \left[\frac{n}{n+1} \right]^{a-2} \left(\frac{\eta}{\mu} + \frac{\lambda Y_n - A_n}{\mu \left(1 + \frac{1}{n} Y_n \right)} \right)$$

where

$$A_n = n \left(\left[1 - \frac{1}{n} \right]^{a-1} - 1 \right).$$

Observe that $\lim_{n \rightarrow \infty} A_n = 1 - a$. The limiting recursion for Y_n is therefore

$$Y_{n+1} - [1 - a] \approx \frac{\eta}{\mu} + \frac{\lambda}{\mu} (Y_n - [1 - a]),$$

and this has the unique fixed point $Y_{\infty} = 1 + \zeta - a$. One can verify that Y_n does indeed converge to Y_{∞} . Thus $a < \zeta$ guarantees convergence.

E PROOF OF PROPOSITION 4

In this Appendix, write (μ_H, λ_H) for $(\mu_{z_H}, \lambda_{z_H})$ and (μ_L, λ_L) for $(\mu_{z_L}, \lambda_{z_L})$.

Let $M_{H,n,t}$ and $M_{L,n,t}$ denote the measures of firms with n blueprints in the high- and low-growth regimes, respectively. As before, $N_t = \sum_{n=1}^{\infty} n (M_{H,n,t} + M_{L,n,t})$ measures the number of blueprints, and νN_t is the flow of blueprints introduced by new firms. The evolution of $M_{H,n,t}$ is determined by the differential equations (21) and (22), modified to include an additional term $-\delta M_{H,n,t}$ on the right-hand side. Note that this term is not scaled by n , since the transition probability from high to low growth is assumed to be independent of size. The differential equations for $M_{L,n,t}$ are (21) and (22), with an additional term $+\delta M_{H,n,t}$ on the right-hand side, and without the term νN_t that

appears in (21). All entering firms are assumed to start out in the high-growth regime. Starting from an initial condition $\{M_{H,n,0}, M_{L,n,0}\}_{n=1}^{\infty}$, the resulting system of differential equations determines the joint distribution of firms by size and growth regime at all times.

Along a balanced growth path, $[M_{H,n,t}, M_{L,n,t}, N_t] = e^{\eta t}[M_{H,n}, M_{L,n}, N]$, where η is the population growth rate. Given (μ_H, λ_H) and (μ_L, λ_L) , the entry parameter ν must be such that the flow of new commodities introduced by new and incumbent firms equals ηN_t . This implies

$$\eta = \nu + \alpha(\mu_H - \lambda_H) + (1 - \alpha)(\mu_L - \lambda_L), \quad (43)$$

where

$$\alpha = \frac{1}{N} \sum_{n=1}^{\infty} n M_{H,n}$$

is the fraction of all blueprints employed by firms in the high-growth phase. The steady state distribution of blueprints by firm size is determined by α and the conditional distributions $Q_{H,n} \propto n M_{H,n}$ and $Q_{L,n} \propto n M_{L,n}$. The differential equations for $M_{H,n,t}$ and $M_{L,n,t}$ imply

$$(\eta + \delta)Q_{H,1} = \lambda_H Q_{H,2} + \frac{\nu}{\alpha} - (\lambda_H + \mu_H)Q_{H,1} \quad (44)$$

and

$$\frac{1}{n}(\eta + \delta)Q_{H,n} = \lambda_H Q_{H,n+1} + \mu_H Q_{H,n-1} - (\lambda_H + \mu_H)Q_{H,n} \quad (45)$$

for all $n - 1 \in \mathbb{N}$, as well as

$$\eta Q_{L,1} = \lambda_L Q_{L,2} - (\lambda_L + \mu_L)Q_{L,1} + \frac{\alpha\delta}{1 - \alpha}Q_{H,1} \quad (46)$$

and

$$\frac{\eta}{n}Q_{L,n} = \lambda_L Q_{L,n+1} + \mu_L Q_{L,n-1} - (\lambda_L + \mu_L)Q_{L,n} + \frac{\alpha\delta}{1 - \alpha} \frac{1}{n}Q_{H,n} \quad (47)$$

for all $n - 1 \in \mathbb{N}$. In addition, $\{Q_{H,n}\}_{n=1}^{\infty}$ and $\{Q_{L,n}\}_{n=1}^{\infty}$ must be probability distributions,

$$\sum_{n=1}^{\infty} Q_{H,n} = \sum_{n=1}^{\infty} Q_{L,n} = 1. \quad (48)$$

The balanced growth distribution of blueprints across firms is now defined by a fraction $\alpha \in (0, 1)$, an entry parameter $\nu > 0$, and a non-negative sequence $\{Q_{H,n}, Q_{L,n}\}_{n=1}^{\infty}$ that satisfy (43)-(48).

Observe that (44)-(45) and (48) exactly match the one-regime conditions (23)-(25) with ν and η replaced by ν/α and $\eta + \delta$, respectively. It must therefore be that $\nu/\alpha =$

$\eta + \delta - (\mu_{\text{H}} - \lambda_{\text{L}})$. Combining this with (43) gives the solution for α and ν ,

$$\alpha = \frac{\eta - (\mu_{\text{L}} - \lambda_{\text{L}})}{\eta + \delta - (\mu_{\text{L}} - \lambda_{\text{L}})}, \quad \nu = \eta + \delta - (\mu_{\text{H}} - \lambda_{\text{H}}).$$

The equilibrium conditions for a balanced growth path ensure $\alpha \in (0, 1)$ and $\nu > 0$. As in the one-regime case, define $\beta_{\text{H},1} = \beta_{\text{L},1} = 0$ and

$$\beta_{\text{H},n+1} = \left(1 + \frac{\eta + \delta + \mu_{\text{H}}n}{\lambda_{\text{H}}n} - \frac{\mu_{\text{H}}\beta_{\text{H},n}}{\lambda_{\text{H}}} \right)^{-1}, \quad \beta_{\text{L},n+1} = \left(1 + \frac{\eta + \mu_{\text{L}}n}{\lambda_{\text{L}}n} - \frac{\mu_{\text{L}}\beta_{\text{L},n}}{\lambda_{\text{L}}} \right)^{-1},$$

for all $n \in \mathbb{N}$. As in Proposition 2, $\beta_{\text{H},n} \uparrow \min\{1, \lambda_{\text{H}}/\mu_{\text{H}}\}$ and $\beta_{\text{L},n} \uparrow \min\{1, \lambda_{\text{L}}/\mu_{\text{L}}\}$. Proposition 2 and the solution for ν/α then imply

$$Q_{\text{H},n} = \frac{\eta + \delta - (\mu_{\text{H}} - \lambda_{\text{H}})}{\mu_{\text{H}}} \sum_{k=n}^{\infty} \frac{1}{\beta_{\text{H},k+1}} \left(\prod_{m=n+1}^{k+1} \beta_{\text{H},m} \right) \prod_{m=2}^{k+1} \frac{\mu_{\text{H}}\beta_{\text{H},m}}{\lambda_{\text{H}}} \quad (49)$$

for all $n \in \mathbb{N}$. Furthermore, (46) and (47) correspond to $Q_{\text{L},n} = \beta_{\text{L},n+1} (Q_{\text{L},n+1} + Z_{\text{L},n+1})$ and

$$Z_{\text{L},n+1} = \left(\frac{\mu_{\text{L}}\beta_{\text{L},n}}{\lambda_{\text{L}}} \right) Z_{\text{L},n} + \frac{1}{\lambda_{\text{L}}} \frac{\alpha\delta}{1 - \alpha} \frac{1}{n} Q_{\text{H},n}$$

for all $n \in \mathbb{N}$, starting from $Z_{\text{L},1} = 0$. This last recursion and its initial condition determine $\{Z_{\text{L},n}\}_{n=1}^{\infty}$. Solving the recursion for $Q_{\text{L},n}$ forward then gives

$$Q_{\text{L},n} = \frac{1}{\mu_{\text{L}}} \frac{\alpha\delta}{1 - \alpha} \sum_{k=n}^{\infty} \frac{1}{\beta_{\text{L},k+1}} \left(\prod_{m=n+1}^{k+1} \beta_{\text{L},m} \right) \sum_{l=1}^k \left(\prod_{m=l+1}^{k+1} \frac{\mu_{\text{L}}\beta_{\text{L},m}}{\lambda_{\text{L}}} \right) \frac{1}{l} Q_{\text{H},l}. \quad (50)$$

The overall firm size distribution is determined by $[\alpha Q_{\text{H},n} + (1 - \alpha)Q_{\text{L},n}]/n$ and (49)-(50).

The rate at which the right tail of this distribution decays is determined by the slowest rate of decay of $Q_{\text{H},n}/n$ and $Q_{\text{L},n}/n$. For the high-growth regime, Proposition 3 shows that the size distribution has a tail index $(\eta + \delta)/(\mu_{\text{H}} - \lambda_{\text{H}})$ if $\mu_{\text{H}} > \lambda_{\text{H}}$. It remains to determine the properties of the right tail of $Q_{\text{L},n}/n$.

The bounds given in Proposition 2 imply that there are positive $\underline{Q} > \underline{Q}$ so that

$$Q_{\text{L},n} \in [\underline{Q}, \overline{Q}] \sum_{k=n}^{\infty} \left(\prod_{m=n+1}^{k+1} \beta_{\text{L},m} \right) \sum_{l=1}^k \frac{1}{l} \prod_{m=2}^{l+1} \frac{\mu_{\text{H}}\beta_{\text{H},m}}{\lambda_{\text{H}}} \prod_{m=l+1}^{k+1} \frac{\mu_{\text{L}}\beta_{\text{L},m}}{\lambda_{\text{L}}}$$

for all $n \in \mathbb{N}$. By replacing $1/l$ with $1/k$ and either keeping only the $l = 1$ term or only the $l = k$ term one obtains the lower bound

$$Q_{\text{L},n} \geq \left(\prod_{m=2}^n \frac{\mu_{\text{L}}\beta_{\text{L},m}}{\lambda_{\text{L}}} \right) \times \underline{Q} \sum_{k=n}^{\infty} \frac{1}{k} \prod_{m=n+1}^{k+1} \beta_{\text{L},m} \prod_{m=n+1}^{k+1} \frac{\mu_{\text{L}}\beta_{\text{L},m}}{\lambda_{\text{L}}},$$

for $I \in \{\text{L}, \text{H}\}$ and some positive $\underline{\underline{Q}}$. Proposition 2 and its proof can be used to argue that the infinite sum on the right-hand side of this inequality is finite and bounded away from zero. Define

$$\sigma_n = \max \left\{ \frac{\mu_{\text{L}} \beta_{\text{L},n}}{\lambda_{\text{L}}}, \frac{\mu_{\text{H}} \beta_{\text{H},n}}{\lambda_{\text{H}}} \right\}.$$

An upper bound for $Q_{\text{L},n}$ is then

$$Q_{\text{L},n} \leq \left(\prod_{m=2}^n \sigma_m \right) \times \underline{\underline{Q}} \sum_{k=n}^{\infty} \left(\sum_{l=1}^k \frac{1}{l} \right) \prod_{m=n+1}^{k+1} \sigma_m$$

for some positive $\underline{\underline{Q}}$. Note that $\sum_{l=1}^k 1/l \leq 1 + \ln(k)$. The infinite sum converges because $\ln(k) \prod_{m=1}^k \sigma_m$ is summable, which in turn follows because $k^\varepsilon \prod_{m=1}^k \sigma_m$ is summable for some small enough $\varepsilon > 0$. This can be shown along the lines of the proof of Proposition 3.

These upper and lower bounds on $Q_{\text{L},n}$ imply that its tail behavior is determined by $\prod_{m=2}^n \sigma_m$. This is geometrically bounded if $\mu_{\text{H}} < \lambda_{\text{H}}$ and $\mu_{\text{L}} < \lambda_{\text{L}}$. If $\mu_{\text{H}} > \lambda_{\text{H}}$ and $\mu_{\text{L}} < \lambda_{\text{L}}$ then the tail index is $(\eta + \delta)/(\mu_{\text{H}} - \lambda_{\text{H}})$; if $\mu_{\text{H}} < \lambda_{\text{H}}$ and $\mu_{\text{L}} > \lambda_{\text{L}}$ then it is $\eta/(\mu_{\text{L}} - \lambda_{\text{L}})$. For the case $\mu_{\text{H}} > \lambda_{\text{H}}$ and $\mu_{\text{L}} > \lambda_{\text{L}}$, define

$$B_{\text{H},n} = n \left(1 - \frac{\mu_{\text{H}} \beta_{\text{H},n}}{\lambda_{\text{H}}} \right), \quad B_{\text{L},n} = n \left(1 - \frac{\mu_{\text{L}} \beta_{\text{L},n}}{\lambda_{\text{L}}} \right).$$

One can use the recursions for $\beta_{\text{H},n}$ and $\beta_{\text{L},n}$ to set up recursions for $B_{\text{H},n}$ and $B_{\text{L},n}$. These show that $B_{\text{H},n} \rightarrow (\eta + \delta)/(\mu_{\text{H}} - \lambda_{\text{H}})$ and $B_{\text{L},n} \rightarrow \eta/(\mu_{\text{L}} - \lambda_{\text{L}})$. Hence

$$\lim_{n \rightarrow \infty} n \left(\frac{\mu_{\text{H}} \beta_{\text{H},n}}{\lambda_{\text{H}}} - \frac{\mu_{\text{L}} \beta_{\text{L},n}}{\lambda_{\text{L}}} \right) = \lim_{n \rightarrow \infty} (B_{\text{L},n} - B_{\text{H},n}) = \frac{\eta}{\mu_{\text{L}} - \lambda_{\text{L}}} - \frac{\eta + \delta}{\mu_{\text{H}} - \lambda_{\text{H}}}.$$

When the two tail indices are not the same, this implies $\sigma_n = \mu_{\text{L}} \beta_{\text{L},n} / \lambda_{\text{L}}$ or $\sigma_n = \mu_{\text{H}} \beta_{\text{H},n} / \lambda_{\text{H}}$ for all but finitely many n , and from this the proposition follows.

F PROOF OF PROPOSITION 5

6.1 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

$$\Pr [Z_k = n] = \sum_{m=1}^{\min\{k,n\}} \binom{k}{m} \binom{n-1}{m-1} (1-\theta)^m \theta^{k-m} (1-\gamma)^m \gamma^{n-m} \quad (51)$$

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.

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

$$(1-\sigma) \sum_{k=1}^{\infty} \sigma^{k-1} \Pr [Z_k = n] = \frac{(1-\sigma)\gamma^n}{\sigma(1-\theta\sigma)} \sum_{m=1}^n \binom{n-1}{m-1} \left[\frac{\sigma(1-\theta)(1-\gamma)}{(1-\theta\sigma)\gamma} \right]^m \quad (52)$$

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

$$(1-\sigma) \sum_{n=N}^{\infty} \sum_{k=1}^{\infty} \sigma^{k-1} \Pr [Z_k = n] = \frac{1-\theta}{1-\theta\sigma} \left[\frac{(1-\sigma)\gamma + \sigma(1-\theta)}{1-\theta\sigma} \right]^{N-1} \quad (53)$$

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

6.2 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 $s_0(a) = \frac{\lambda}{\mu} \gamma(a)$. Conditional on survival, its size distribution is the geometric size distribution $s_n(a)/[1-s_0(a)] = [1-\gamma(a)]\gamma^{n-1}(a)$. This can be verified directly by checking (32)-(33). 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 (51) 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 (31)-(32), as described in the text.

The infinite sums needed in (34) and (35) follow from (52) and (53). Age densities (distributions) can then be computed using a univariate (bivariate) numerical integration.

REFERENCES

- [1] Arkolakis, Costas, “Market Access Costs and the New Consumers Margin in International Trade,” working paper, University of Minnesota (2006).
- [2] Arrington, Catherine, “Statistics on U.S. Businesses—Microdata and Tables,” Small Business Administration (1998).
- [3] Atkeson, Andrew, and Patrick Kehoe, “Modelling and Measuring Organization Capital,” *Journal of Political Economy*, vol. 113, no. 5 (2005), 1026-1051.
- [4] Axtell, Robert L., “Zipf Distribution of U.S. Firm Sizes,” *Science*, CCXCIII (2001), 1818-1820.
- [5] Beckmann, M.J., “City Hierarchies and the Distribution of City Size,” *Economic Development and Cultural Change*, vol. 6, no. 3 (1958), 243-248.
- [6] Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, “Multi-Product Firms and Product Switching,” NBER working paper no. 12293 (2006).
- [7] Boldrin, Michele, and David K. Levine, “Perfectly Competitive Innovation,” working paper (1999, 2006).
- [8] Chandler, Alfred D., Jr., *Scale and Scope—The Dynamics of Industrial Capitalism*, Harvard University Press (1994).
- [9] Chatterjee, Satyajit and Estreban Rossi-Hansberg, “Spin-offs and the Market for Ideas (2006).
- [10] Davis, Steven J., John Haltiwanger, Ron Jarmin and Javier Miranda, “Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms,” *NBER Macroeconomics Annual 2006, Volume 21*, edited by Daron Acemoglu, Kenneth Rogoff and Michael Woodford, 107-156.

- [11] Dixit, Avinash K., and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, LXXVII (1977), 297-308.
- [12] Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, "The Growth and Failure of U.S. Manufacturing Plants," *Quarterly Journal of Economics*, CIV (1989), 671-698.
- [13] Eaton, J. and S. Kortum, "Technology, Geography and Trade," *Econometrica*, vol. 70, no. 5 (2002), 1741-1779.
- [14] Evans, D.S., "The Relationship between Firm Growth, Size and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics*, vol. XXXV, no. 4, 567-581, 1987.
- [15] Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, CXIV (1999), 739-367.
- [16] Gabaix, Xavier and Augustin Landier, "Why Has CEO Pay Increased so Much," *Quarterly Journal of Economics*, vol. 123, no. 1 (2008), 49-100.
- [17] Hall, B. H., "The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector," *Journal of Industrial Economics*, vol. XXXV, no. 4, 583-605, 1987.
- [18] Holmes, Thomas J., "The Diffusion of Wal-Mart and Economies of Density," University of Minnesota working paper (2006).
- [19] Holmes, Thomas J. and James A. Schmitz, Jr., "On the Turnover of Business Firms and Business Managers," *Journal of Political Economy*, vol. 103, no. 5 (1995), 1005-1038.
- [20] Hopenhayn, Hugo, "Entry, Exit and Firm Dynamics in Long Run Equilibrium," *Econometrica*, vol. 60, no. 5 (1992), 1127-1150.
- [21] Hsu, Wen-Tai, "Central Place Theory and Zipf's Law," working paper, University of Minnesota (2007).
- [22] Hymer, Stephen and Peter Pashigian, "Firm Size and Rate of Growth," *Journal of Political Economy*, vol. 70, no. 6 (1962), 556-569.

- [23] Ijiri, Yuji, and Herbert A. Simon, "Business Firm Growth and Size," *American Economic Review*, LIV (1964), 77-89.
- [24] Jackson, Matthew O., "The Economics of Social Networks," in *Proceedings of the 9th World Congress of the Econometric Society*, edited by Richard Blundell, Witney Newey, and Torsten Persson, 1-56, Cambridge University Press, Cambridge MA (2006).
- [25] Jackson, Matthew O. and Brian W. Rogers, "Meeting Strangers and Friends of Friends: How Random Are Social Networks," *American Economic Review*, vol. 97, no. 3 (2007), 890-915.
- [26] Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, C (1982), 649-670.
- [27] Jovanovic, Boyan and Peter L. Rousseau, "A Q-Theory of Mergers," *American Economic Review, Papers and Proceedings*, vol. 92, no. 2 (2002), 198-204.
- [28] Kendall, David G., "On the Generalized "Birth-and-Death" Process," *The Annals of Mathematical Statistics*, vol. 19, no. 1 (1948), 1-15.
- [29] Klette, Tor Jakob, and Samuel S. Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, vol. 112, no. 5 (2004), 986-1018.
- [30] Lentz, Rasmus and Dale T. Mortensen, "An Empirical Model of Growth Through Product Innovation" University of Wisconsin and Northwestern University (2006).
- [31] Lucas, Robert E., Jr., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, vol. 9, no. 2 (1978), 508-523.
- [32] Luttmer, Erzo G.J., "Consumer Search and Firm Growth," Federal Reserve Bank of Minneapolis working paper no. 645 (2006).
- [33] Luttmer, Erzo G.J., "Selection, Growth, and the Size Distribution of Firms, *Quarterly Journal of Economics*, 2007, vol. 122, no. 3 (2007), 1103-1144.
- [34] McFadden, D.P., "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, edited by P. Zarembka: New York, Academic Press (1974).

- [35] Rosen, S., "Substitution and Division of Labour," *Economica*, vol. 45, no. 179 (1978), 235-250.
- [36] Rossi-Hansberg, Esteban and Mark L.J. Wright, "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review*, vol. 97, no. 5 (2007), 1639-1666.
- [37] Seker, Murat, "A Structural Model of Establishment and Industry Evolution: Evidence from Chile," working paper, University of Minnesota (2007).
- [38] Simon, Herbert A., "On a Class of Skew Distribution Functions," *Biometrika*, vol. 42, no. 3/4 (1955), 425-440.
- [39] Simon, Herbert A., and Charles P. Bonini, "The Size Distribution of Business Firms," *American Economic Review*, XCVIII (1958), 607-617.
- [40] Stanley, R. Michael, L.A.N. Amaral, S.V. Buldyrev, S. Harlin, H. Leschorn, P. Maass, M.A. Salinger, H.E. Stanley, "Scaling Behavior in the Growth of Companies," *Nature* 319 (1996), 804-806.
- [41] Sutton, John, "Gibrat's Legacy," *Journal of Economic Literature*, vol. XXXV (1997), 40-59.
- [42] Sutton, John, "The Variance of Firm Growth Rates: The 'Scaling' Puzzle," *Physica A*, CCCXII (2002), 577-590.
- [43] Sutton, John, "Market Share Dynamics and the 'Persistence of Leadership' Debate," *American Economic Review*, vol. 97, no. 1 (2007), 222-241.
- [44] Terviö, Marko, "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, forthcoming (2007).
- [45] Yule, G. Udny, "A Mathematical Theory of Evolution Based on the Conclusions of Dr. J.C. Willis, F.R.S.," *Philosophical Transactions of the Royal Society of London, Series B, Containing Papers of a Biological Character*, vol. 213 (1925), 21-87.