



OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Selection, Growth, and the Size Distribution of Firms

Author(s): Erzo G. J. Luttmer

Source: *The Quarterly Journal of Economics*, Vol. 122, No. 3 (Aug., 2007), pp. 1103-1144

Published by: Oxford University Press

Stable URL: <http://www.jstor.org/stable/25098869>

Accessed: 17-11-2016 17:58 UTC

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.org/stable/25098869?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>



Oxford University Press is collaborating with JSTOR to digitize, preserve and extend access to *The Quarterly Journal of Economics*

SELECTION, GROWTH, AND THE SIZE DISTRIBUTION OF FIRMS*

ERZO G. J. LUTTMER

This paper describes an analytically tractable model of balanced growth that is consistent with the observed size distribution of firms. Growth is the result of idiosyncratic firm productivity improvements, selection of successful firms, and imitation by entrants. Selection tends to improve aggregate productivity at a fast rate if entry and imitation are easy. The empirical phenomenon of Zipf's law can be interpreted to mean that entry costs are high or that imitation is difficult, or both. The small size of entrants indicates that imitation must be difficult. A calibration based on U. S. data suggests that about half of output growth can be attributed to selection. But the implied variance of the combined preference and technology shocks is puzzlingly high.

I. INTRODUCTION

This paper presents an analytically tractable model of growth resulting from firm-specific preference and technology shocks, selective survival of successful firms, and imitation by entering firms. The model generates balanced growth and is consistent with salient features of the firm size distribution. As many have noted, the size distribution of firms exhibits a striking pattern. Using 1997 data from the U. S. Census, Axtell [2001] finds that the log right tail probabilities of this distribution, with firm size measured by the log of employment, are on a virtual straight line with a slope of -1.06 . Figure I shows the data for 2002, together with a curve generated by a version of the model presented in this paper, as well as the maximum likelihood estimate of a lognormal distribution. A straight line fitted using all size categories with at least five employees has a slope of -1.06 . This evidence suggests that the firm size distribution, with firm size measured by employment S , is well approximated over much of its range by a Pareto distribution with right tail probabilities of the form $1/S^\zeta$, with a tail index ζ around 1.06 .¹

* 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. I thank Michele Boldrin, Jonathan Eaton, Xavier Gabaix, Thomas J. Holmes, Samuel S. Kortum, and Robert E. Lucas, Jr., for helpful discussions based on earlier versions of this paper. Two referees provided valuable input. The usual disclaimer applies. A technical appendix is available at www.luttmer.org.

1. The data shown in Figure I summarize a population of 5,697,759 U. S. firms in 2002. The largest size category, that of 10,000 employees and over, still contains 913 firms. There is a size category of zero employees (in March of 2002),

© 2007 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

The Quarterly Journal of Economics, August 2007

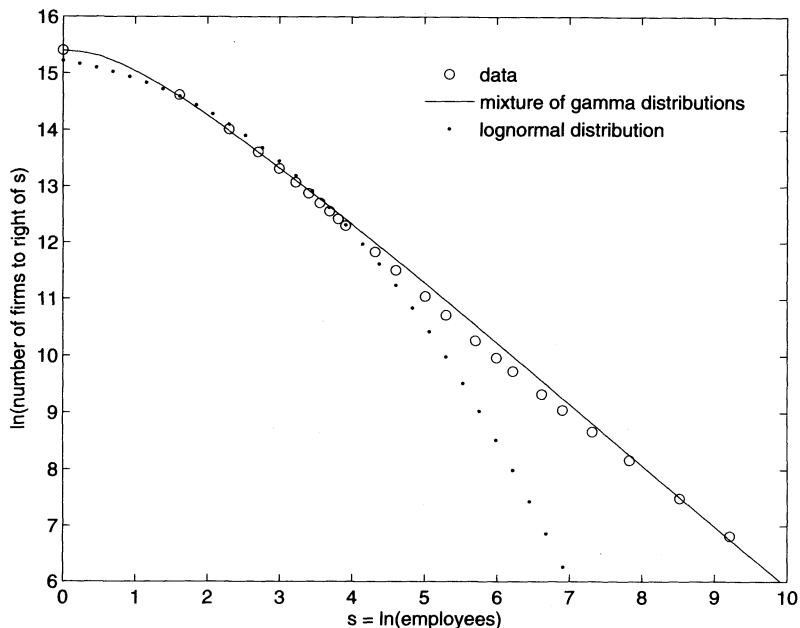


FIGURE I
Size Distribution of U. S. Firms in 2002

The remarkable fit of this distribution has been documented and interpreted before, perhaps most notably by Simon and Bonini [1958], Steindl [1965], and Ijiri and Simon [1964]. As far back as Gibrat [1931], researchers have related the shape of the observed size distribution to models of firm entry, random growth, and exit. The mechanism described in this paper is most like the one proposed for the city size distribution by Gabaix [1999].² In contrast to this literature, this paper explains the observed firm size distributions in terms of primitives such as entry and fixed

accounting for 770,041 firms, that is not shown. The data are originally from the U. S. Census Bureau and were obtained from the Small Business Administration Internet site and from the Statistics of U. S. Businesses site of the U. S. Census Bureau (the size categories 5,000–9,999 and 10,000 and over). The fitted curve represents a mixture of gamma distributions, as discussed in Section VI.C.

2. Sutton [1997] surveys the literature on firm size and Gibrat's law: firm growth is independent of size. Gabaix [1999] contains extensive discussions of the literature on probability models that give rise to Pareto distributions and their application in economics. Gabaix and Ioannides [2003] survey the literature on Zipf's law for cities.

costs, and the ease with which firms can imitate. The explanation is set in the context of a general equilibrium model, and this allows one to predict the effects of changes in various barriers to entry on the level and the growth rate of aggregate output. The model can also be extended in a tractable way to accommodate more extensive forms of heterogeneity [Luttmer 2004], making it a potentially useful tool for empirical research on the relation between firm heterogeneity and aggregate productivity.

Firms in this paper are monopolistic competitors producing differentiated goods, as in Dixit and Stiglitz [1977], using a linear technology. There is an entry cost for new firms, and it takes a fixed cost per unit of time to continue an existing firm. A typical firm is subject to shocks to both productivity and the demand for its differentiated good. These shocks are firm-specific and permanent.³ A stationary firm size distribution arises if the average rate at which these shocks improve the profitability of incumbent firms is not too high relative to the rate at which the technology available to potential entrants improves over time.

One version of this economy is a model of technology adoption in which the technologies available to potential entrants improve at an exogenous rate. This rate determines the growth rate of the economy. If there is not too much heterogeneity among entrants, then the equilibrium size distribution is well approximated, over much of its range, by a Pareto distribution. A tail index ζ slightly above 1 arises if the technologies available to entrants improve at a rate that is only slightly above the rate at which the technologies of incumbents improve. In this economy, a proportional increase in entry and fixed costs lowers the level of aggregate output by reducing the number of firms and thereby the variety of goods produced. This is analogous to results for static economies in Krugman [1979]. The shape of the size distribution is not affected by proportional changes in entry and fixed costs. A reduction in the entry cost alone does change the shape of the size distribution, although not its tail index. Lower entry costs lead to

3. See Melitz [2003] for a related model that features firm heterogeneity, monopolistic competition, together with entry and fixed costs. Much of what follows can be shown also in an economy with perfectly competitive final goods markets, decreasing returns at the firm level and firm-specific technology shocks. This would give rise to an economy similar to Lucas [1978], Hopenhayn [1992], Atkeson and Kehoe [2005], and Hellwig and Irmen [2001]. Most data sets show a lot of heterogeneity across firms, even within narrowly defined industries. An advantage of the monopolistic competition formulation is that shocks to the demands for differentiated goods can be a source of firm heterogeneity, above and beyond firm-specific technology shocks.

more firms and more variety, but the positive effect of this on the level of output is weakened by the fact that more inefficient firms will enter and survive.⁴

A second version of this economy is a model of endogenous growth in which entering firms can imperfectly imitate incumbent firms. This makes the tail index ζ endogenous. A potential entrant can pay an entry cost to sample at random from the population of incumbent firms. The entrant can then attempt to imitate the incumbent drawn from the population by introducing a new good with an initial productivity and market size that are scaled down relative to the productivity and market size of the incumbent. This spillover ensures that the technologies available to potential entrants are never so far behind those of incumbent firms that entry of new firms is not feasible. The economy has a continuum of stationary size distributions that are consistent with balanced growth. One possibility is that the log of firm size follows a gamma distribution. All possible size distributions have a tail similar to that of a Pareto distribution, with an analogous tail index ζ that must be slightly above 1 to fit the data shown in Figure I. The main result for this economy is that ζ converges to 1 from above as the cost of entry becomes large relative to the fixed cost of operating a firm, and as the extent to which new entrants lag behind incumbents in terms of productivity and market size, becomes large.

To see why the asymptote $\zeta = 1$ arises, note that the mean of a distribution with right tail probabilities of the order $1/S^\zeta$ grows without bound as ζ approaches 1 from above. Firm profitability is tied to size, and the fact that potential entrants attempt to imitate a randomly sampled incumbent ties the expected gains from entry to the average size of incumbents. In equilibrium, high entry costs must be compensated for by high expected gains from entry. Thus, the average incumbent must be large, and especially so if entrants lag far behind incumbents in terms of productivity and market size.

As in the version with exogenous growth, a proportional reduction in entry and fixed costs increases the level of output in this economy. The effect of lowering entry costs alone is to lower the average size and profitability of firms. This is achieved in

4. See Parente and Prescott [1999] for an alternative model of technology adoption in which lowering barriers to entry can have large positive effects on the level of output.

equilibrium by an increase in the turnover rate of firms. In turn, this speeds up the selection mechanism by which aggregate productivity improves over time. As a result, the growth rate of the economy increases. A reduction in barriers to entry will, over time, have large effects on output when entrants can imitate incumbents. This is in sharp contrast to the level effect that arises when the technologies available to entrants are exogenous.

The firm size distribution, together with data on the size of entering firms and the rate at which new firms enter can be used to infer the parameters of the firm growth process. These parameters imply a decomposition of output growth into components due to within-firm technological progress and selection. U. S. data suggest that about half of output growth can be attributed to selection. The parameter estimates also produce predictions for the hazard rate with which firms exit, and these are in line with observed survivor functions. However, the variance of firm growth rates is higher than suggested by the return variance of the typical firm traded in U. S. stock markets.

I.A. Related Literature

Incumbent firms in this paper are engaged in a form of learning-by-doing, and imitation by entering firms creates an externality, two features of growth emphasized by Arrow [1962].⁵ Following Romer [1990], Grossman and Helpman [1991], and Aghion and Howitt [1992], technological progress is embodied in firms, and firms have some market power. As in Romer [1990], this takes the form of monopolistic competition.⁶ The current paper differs in two important respects from Romer [1990]. First, firms experience idiosyncratic permanent shocks to their technologies and to the demands for their differentiated commodities. This introduces selection as a mechanism by which the economy-wide distribution of productivity improves over time. Random growth and selection are crucial for matching the observed firm size distribution. Second, the mechanism that allows potential entrants to make use of the existing stock of ideas is made explicit. This yields an economic interpretation of the size distri-

5. The more recent literature making use of these features includes Boldrin and Scheinkman [1988], Lucas [1988], Stokey [1988], and Young [1991].

6. Jones and Manuelli [1990] and Boldrin and Levine [2000] construct models of endogenous growth that do not rely on imperfect competition or externalities.

bution shown in Figure I: imitation is imperfect and entry must be costly.⁷

In Jovanovic [1982], the effects of selection on the evolution of an industry eventually die out because firms are not subject to ongoing technology shocks. In Hopenhayn [1992], the industry equilibrium is stationary, but there is no reason for the implied size distribution to look like the one displayed in Figure I. In this paper, all shocks to preferences and technology are permanent. Stationarity of the cross-sectional size distribution is a consequence of the spillover that relates the productivity of entrants to the distribution of productivity among incumbents.

Gabaix [1999] shows how a geometric Brownian motion with a reflecting barrier gives rise to a power law and shows the precise circumstances under which this will lead to Zipf's law. He uses this to construct a model of cities that can account for the heavy right tail of the city size distribution. In the presence of entry and fixed costs, the process of firm entry and exit does not lead to a reflecting barrier, but to a "return process" according to which firms exit below some barrier and enter at a point above this barrier. The two processes are closely related, and the limiting argument used by Gabaix [1999] will be discussed later. Essentially the same return process as used in the technology adoption part of this paper also arises in Miao [2005], who considers a model of industry equilibrium and debt-financing in which default triggers exit.

Based on a data set that includes not only large cities, Eeckhout [2004] has argued that the size distribution of cities or "places" is approximately lognormal rather than Pareto. The maximum-likelihood estimate shown in Figure I shows that the lognormal distribution is greatly at odds with the observed size distribution of firms. Just like the lognormal distribution, the gamma distributions generated in this paper have a mode that exceeds the minimum firm size. In contrast to the lognormal, these gamma distributions can also match the heavy right tail of the firm size distribution.

The economy described here has many elements in common

7. Jovanovic [1982] emphasizes the role of selection in the evolution of an industry. Nelson and Winter [1982] relate selection, imitation, and growth, but their model is not analytically tractable. Jovanovic and MacDonald [1994] consider industry growth with very general forms of imitation. Other models of imitation and growth include Segerstrom [1991], Aghion, Harris, Howitt, and Vickers [2001], and Eeckhout and Jovanovic [2002]. Barro and Sala-i-Martin [2004] present models of growth that rely on cross-country imitation.

with Klette and Kortum [2004], who build on Grossman and Helpman [1991] to construct a quality ladder model in which firm growth is the result of research and development choices made by firms. Every good produced by a firm can give rise to a new good or can be lost to a competitor following exponentially distributed waiting times. As a result, the underlying building block of the model is a birth and death process for the number of goods produced by a firm. In this paper it is a geometric Brownian motion that represents the state of consumer tastes and firm productivity. For both processes, mean growth rates are independent of size. In the case of the geometric Brownian motion, the same is true for the variance of firm growth rates. In the case of the birth and death process, averaging across goods implies that the variance is inversely proportional to size. The resulting size distribution is the logarithmic series distribution. This distribution is highly skewed, but a plot as in Figure I generates a curve that is concave and does not asymptote to a straight line for large firm sizes. The right tail of the distribution is too thin.

Rossi-Hansberg and Wright [2004] solve for the firm size distribution in an economy with many industries and many identical firms in each industry. Firms face a fixed cost in every period and operate Cobb–Douglas technologies that exhibit decreasing returns. Human capital is industry specific, and the number and size of firms in a particular industry at a point in time is determined by a static free-entry condition. Because of this static free-entry condition, it does not matter which of the infinitesimal firms in an industry exit when net exit from a particular industry is required. As a result, the model has no determinate implications for the dependence of firm exit rates on age, or for the joint age-size distribution of firms. In equilibrium, the industry-specific human capital stock exhibits mean reversion, and this generates a stationary firm size distribution. If shocks to the human capital accumulation technology are lognormal, then the size distribution is lognormal. As shown in Figure I, the lognormal distribution has many fewer large firms than are observed in the data.

I.B. Outline of the Paper

The model of technology adoption is set up in Section II. The size distribution is characterized in Section III and the balanced growth path is determined in Section IV. Imitation is introduced in Section V, and the relations between entry costs, the size

distribution, and the growth rate of the economy are described. Section VI presents calibrations, allowing for multiple industries with different cost structures and growth rates. Concluding remarks are in Section VII.

II. TECHNOLOGY ADOPTION

II.A. Consumers

Time is continuous and indexed by t . There is a continuum of consumers alive at any point in time. The population size at time t is $He^{\eta t}$, and the population growth rate η is nonnegative. During their lifetimes, consumers supply one unit of labor at every point in time. There is a representative consumer with preferences over rates of dynastic consumption $\{C_t\}_{t \geq 0}$ of a composite good, defined by the utility function,

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

The discount rate ρ and the intertemporal elasticity of substitution $1/\gamma$ are positive. The composite good is made up of a continuum of differentiated commodities. Preferences over these commodities are additively separable with weights that define the type of a commodity. This implies that all commodities of the same type and trading at the same price are consumed at the same rate. Let $c_t(u, p)$ be consumption at time t of a commodity of type u that trades at a price p . In equilibrium, there will be a measure M_t of commodities that are available at time t , defined on the set of commodity types and prices. The composite good is a version of the one specified in Dixit and Stiglitz [1977]. For some $\beta \in (0, 1)$,

$$(1) \quad C_t = \left[\int u^{1-\beta} c_t^\beta(u, p) dM_t(u, p) \right]^{1/\beta}.$$

The type u of a commodity can be viewed as measure of its quality. The level of $c_t(u, p)$ is chosen to minimize the cost of acquiring C_t . This implies that

$$(2) \quad p c_t(u, p) = P_t(u C_t)^{1-\beta} c_t^\beta(u, p),$$

where P_t is the price index:

$$(3) \quad P_t = \left[\int u p^{-\beta/(1-\beta)} dM_t(u, p) \right]^{-(1-\beta)/\beta}.$$

The price elasticity of the demand for commodity (u, p) is $-1/(1 - \beta)$, and the implied expenditure share is $u(p/P_t)^{-\beta/(1-\beta)}$.

The representative consumer faces a standard present-value budget constraint. The consumer's wealth consists of claims to firms and labor income. Along the balanced growth path constructed below, per capita consumption and real wages grow at a common rate κ . The paths of per capita consumption and real wages are denoted by $C_t e^{-\eta t} = C e^{\kappa t}$ and $w_t = w e^{\kappa t}$. When the composite good is used as the numeraire, the interest rate is constant and given by $r = \rho + \gamma \kappa$. The following assumption ensures that the present value of aggregate consumption and labor income is finite.

ASSUMPTION 1. The growth rates η and κ satisfy $\eta \geq 0$ and $\rho + \gamma \kappa > \kappa + \eta$.

This assumption implies that $\rho > (1 - \gamma)\kappa$, and thus utility is finite.

II.B. Firms

A firm is defined by its unique access to a technology for producing a particular differentiated commodity. At age a , a firm that was set up at time t uses $L_{t,a}$ units of labor to produce $z_{t,a} L_{t,a}$ units of a differentiated commodity of quality $u_{t,a}$. Given a price $p_{t,a}$, the revenues of the firm are given by $R_{t,a} = p_{t,a} z_{t,a} L_{t,a} / P_t$, in units of the composite good. The demand function for type- $u_{t,a}$ commodities implies that these revenues can be written as

$$(4) \quad R_{t,a} = C_{t+a}^{1-\beta} (Z_{t,a} L_{t,a})^\beta,$$

where $Z_{t,a} = (u_{t,a}^{1-\beta} z_{t,a}^\beta)^{1/\beta}$ combines the state of preferences and technology. Firm revenues vary with aggregate consumption, the weight $u_{t,a}$ of its output in the utility function, and its productivity level $z_{t,a}$. With some abuse of terminology, the combination of quality and quantity measured by $Z_{t,a}$ will be referred to simply as productivity. The productivities $Z_{t,a}$ are assumed to evolve independently across firms, according to

$$(5) \quad Z_{t,a} = Z \exp(\theta_B t + \theta_I a + \sigma_Z W_{t,a}),$$

where $\{W_{t,a}\}_{a \geq 0}$ is a standard Brownian motion and Z is an

initial condition.⁸ Note that $Z_{t,0} = Ze^{\theta_E t}$ is the initial productivity of a new firm at time t . Thus θ_E is the rate at which the productivity of entering firms grows over time. The trend of log productivity for incumbent firms is determined by θ_I . The difference between θ_E and θ_I is a key determinant of the firm size distribution. In Section V, θ_E will be made endogenous.

An existing firm can be continued only at a cost equal to λ_F units of labor per unit of time. The firm must exit if this fixed cost is not paid, and exit is irreversible. One interpretation is that it is costly to preserve the information accumulated as a result of past firm-specific shocks to preferences and technology, and that this information is lost as soon as the required costs are not incurred.⁹ Measured in units of the composite good, the value $V_t[Z]$ at time t of a firm with initial productivity $Ze^{\theta_E t}$ is given by

$$V_t[Z] = \max_{L,\tau} E_t \left[\int_0^\tau e^{-ra} (R_{t+a} - w_{t+a} [L_{t,a} + \lambda_F]) da \right].$$

The maximization is subject to (4) and (5) and subject to the restriction that production and exit decisions only depend on the available information.

The aggregate supply of labor grows at a rate η , and every firm must use at least λ_F units of labor to stay in business. Along the balanced growth path, the number of firms grows at the rate η . Entry and exit generates time- t cross-sectional distributions of labor inputs $L_{t-a,a}$ and productivities relative to trend $Z_{t-a,a}e^{-\theta_E t}$ that are time invariant. Since the number of firms grows at a rate η , the growth rate κ of per capita consumption must also be the growth rate of average revenues per firm. Together with (4) this gives

$$(6) \quad \kappa = \theta_E + \left(\frac{1 - \beta}{\beta} \right) \eta.$$

Population growth implies growth in the number of differentiated commodities. This adds to the growth rate θ_E of productivity, with

8. This productivity process will result, for example, if both $u_{t,a}$ and $z_{t,a}$ are geometric Brownian motions.

9. Atkeson and Kehoe [2005] assume perfect competition together with decreasing returns to variable inputs and interpret λ_F as the cost of a managerial fixed factor, along the lines of Lucas [1978]. Much of what follows continues to hold for such an alternative model.

a slope that is large when substitution between these commodities is difficult.

Production Decisions. Firms choose variable labor to maximize variable profits $R_{t,a} - w_{t+a}L_{t,a}$, subject to (4). The optimal choice is

$$(7) \quad \begin{bmatrix} R_{t,a} \\ w_{t+a}L_{t,a} \end{bmatrix} = \begin{bmatrix} 1 \\ \beta \end{bmatrix} \left(\frac{\beta Z_{t,a}}{w_{t+a}} \right)^{\beta/(1-\beta)} C_{t+a}.$$

Together with (5) and (6) this implies that, along the balanced growth path, labor and revenues measured in units of labor do not depend on calendar time. In particular, the revenues net of fixed and variable costs can be written as

$$R_{t,a} - w_{t+a}(L_{t,a} + \lambda_F) = w_{t+a}\lambda_F(e^{s_a} - 1),$$

where s_a equals

$$(8) \quad s_a = S[Z] + \frac{\beta}{1-\beta} \left[\ln \left(\frac{Z_{t,a}}{Z_{t,0}} \right) - \theta_E a \right],$$

and where $S[Z]$ is defined by

$$(9) \quad e^{S[Z]} = \frac{1-\beta}{\lambda_F} \frac{C}{w} \left(\frac{\beta Z}{w} \right)^{\beta/(1-\beta)}.$$

Both revenues and variable labor inputs are proportional to $w_{t+a}\lambda_F e^{s_a}$. The variable s_a can thus be viewed as a measure of firm size relative to fixed costs. If $s_a = 0$, then variable revenues just cover fixed costs. It follows from (5) and (8) that firm size evolves with age according to $ds_a = \mu da + \sigma dW_{t,a}$, where

$$(10) \quad \begin{bmatrix} \mu \\ \sigma \end{bmatrix} = \frac{\beta}{1-\beta} \begin{bmatrix} \theta_I - \theta_E \\ \sigma_Z \end{bmatrix}.$$

Firm size has a negative drift when productivity inside the firm is expected to grow more slowly than the productivity of new entrants. Note that the differences in these growth rates and the variance of productivity shocks are greatly magnified when the differentiated goods are close substitutes.

The function $S[Z]$ defined in (9) plays an important role in the rest of the paper. Along the balanced growth path, where (6) holds, it relates the de-trended productivity of any firm to its size. More precisely, $e^{S[Z]}$ is the size of any firm with productivity $Z e^{\theta_E t}$ at time t , relative to its fixed costs at time t . In particular, it is the

size relative to fixed costs of a new firm entering with a de-trended initial productivity Z .

The Exit Decision. The presence of fixed costs implies a minimum size. Firms with very low productivity choose to exit since they face only a small probability of ever recovering the fixed costs required to continue the firm. The value of a firm of size s relative to its current fixed costs is

$$V(s) = \max_{\tau} E \left[\int_0^{\tau} e^{-(r-\kappa)a} (e^{s_a} - 1) da \mid s_0 = s \right].$$

The value of a firm entering at time t with initial productivity Z is equal to $V_t[Z] = w_t \lambda_F V(S[Z])$. This depends on the level of wages directly via w_t and indirectly via $S[Z]$.

ASSUMPTION 2. Preference and technology parameters satisfy $\rho + \gamma\kappa > \kappa + \mu + \frac{1}{2}\sigma^2$.

Assumption 1 implies that $r > \kappa$, and, thus, the fixed cost of operating a firm forever is finite. Assumption 2 means that $r > \kappa + \mu + \sigma^2/2$, and this implies that the revenues of such a policy are also finite. Together, these assumptions are sufficient to ensure that the value of a firm is finite. The value function $V(s)$ must satisfy the following Bellman equation in the range of s where a firm is not shut down:

$$rV(s) = \kappa V(s) + \mathcal{A}V(s) + e^s - 1,$$

where $\mathcal{A}V(s) = \mu DV(s) + \sigma^2 D^2 V(s)/2$ is the drift of $V(s)$. The return to owning a firm consists of a capital gain $\kappa + \mathcal{A}V(s)/V(s)$ and a dividend yield $(e^s - 1)/V(s)$. It is optimal to shut down a firm when its size s falls below some threshold b . Given that the firm is shut down at b , it must be that the value of a firm is zero at that point. This implies the boundary condition $V(b) = 0$. The optimal threshold must be such that V is differentiable at b , and so $DV(b) = 0$. A further boundary condition follows from the fact that the value function cannot exceed the value of a firm that operates without fixed costs. This implies that $V(s)$ must lie below $e^s/(r - [\kappa + \mu + \sigma^2/2])$.

With these boundary conditions, the Bellman equation has only one solution¹⁰:

$$(11) \quad V(s) = \frac{1}{r - \kappa} \left(\frac{\xi}{1 + \xi} \right) \left(e^{s-b} - 1 - \frac{1 - e^{-\xi(s-b)}}{\xi} \right),$$

for $s \geq b$ and $V(s) = 0$ otherwise. The exit barrier b is determined by

$$(12) \quad e^b = \left(\frac{\xi}{1 + \xi} \right) \left(1 - \frac{\mu + \sigma^2/2}{r - \kappa} \right), \quad \xi = \frac{\mu}{\sigma^2} + \sqrt{\left(\frac{\mu}{\sigma^2} \right)^2 + \frac{r - \kappa}{\sigma^2/2}}.$$

Assumptions 1 and 2 imply that $\xi > 0$, and that b is well defined. As expected, $V(s)$ is strictly increasing on (b, ∞) . It will be useful to note that, for any fixed x , $V(x + b)$ is increasing in ξ and $V(x + b)$ goes to zero as ξ goes to zero. The latter will happen when μ becomes large and negative. If the productivity of new entrants grows very quickly, then the value of being an incumbent at any given distance x away from the exit barrier will be very small.

Entry. New firms can be set up at a cost that is linear in the entry rate. Entry at a rate of l firms per unit of time costs $\lambda_E l$ units of labor per unit of time. Entry results in a draw of Z from a distribution J . At time t , a draw Z yields an initial productivity $Z e^{\theta_E t}$ and thus an initial size $S[Z]$. Along the balanced growth path, entry takes place at all times. This means that the profits from entry must be zero:

$$(13) \quad \lambda_E = \lambda_F \int V(S[Z]) dJ(Z).$$

The distribution J is taken to be exogenous until imitation is introduced in Section V. The only assumption needed here is that the implied value of entry is finite.

ASSUMPTION 3. The initial productivity distribution J satisfies $\int Z^{\beta/(1-\beta)} dJ(Z) < \infty$.

The value of entry depends on steady-state wages and aggregate consumption via $S[Z]$. Recall from (9) that $S[Z]$ is proportional to $(C/w)/w^{\beta/(1-\beta)}$. The returns to entry can therefore be made arbi-

10. See Dixit and Pindyck [1994] for a detailed treatment of closely related stopping problems.

trarily small or large by taking $(C/w)/w^{\beta/(1-\beta)}$ to be small or large, respectively. Thus, the zero-profit condition (13) implies a unique equilibrium value for $(C/w)/w^{\beta/(1-\beta)}$, and therefore also for $S[Z]$. It is not difficult to see that $S[Z]$ is increasing in λ_E . In equilibrium, the initial size and productivity of firms must be high when entry is costly.

III. THE DISTRIBUTION OF FIRM CHARACTERISTICS

There is a continuum of infinitesimal firms. The underlying stochastic structure is assumed to be such that probability distributions for individual firm size can be interpreted as cross-sectional size distributions for the whole continuum of firms.

Along the balanced growth path to be constructed, there is a time-invariant cross-sectional distribution of firm size. Firms enter and exit at constant aggregate rates in such a way that the aggregate measure of firms expands at the rate η . A time-invariant size distribution will result if η is positive, or if η is zero and μ is negative. In any equilibrium, the distribution of firm size, measured by e^s , must also have a finite mean. The following assumption will turn out to be necessary and sufficient for this to be the case, given that η is nonnegative.

ASSUMPTION 4. The productivity parameters satisfy $\eta > \mu + \frac{1}{2} \sigma^2$.

Note that $\mu + \sigma^2/2$ is the drift of the size variable e^{s_a} . Thus, Assumption 4 means that the size of a typical incumbent firm is not expected to grow faster than the population growth rate. If η is zero, then μ must be negative, but otherwise it can be positive.

Although age does not directly affect firm behavior, it will be useful to include age with size as a state variable. Age increases deterministically with a unit drift, and size has drift μ and diffusion coefficient σ . The measure of firms, defined on the set of possible ages a and firm sizes s , grows at a rate η . The density of this measure at date t can be written as $m(a,s)Ie^{\eta t}$, where $Ie^{\eta t}$ is the number of new firms attempting to enter per unit of time. The market clearing conditions that will determine the balanced growth path are linear in m , and this makes it convenient not to normalize m to be a probability density. The density $m(a,s)Ie^{\eta t}$, viewed as a function of the state (a,s) and time t , must satisfy the

Kolmogorov forward equation.¹¹ The resulting partial differential equation for m is given by

$$(14) \quad D_a m(a, s) = -\eta m(a, s) - \mu D_s m(a, s) + \frac{1}{2} \sigma^2 D_{ss} m(a, s)$$

for all $a > 0$ and $s > b$. The first term on the right-hand side of (14) reflects the fact that the measure of firms grows over time. The remaining two terms describe how $m(a, s)$ evolves as a result of stochastic changes in the sizes of individual firms.

Firms use at least λ_F units of labor, and so the measure of firms has to be finite in any equilibrium. As age goes to zero, the size distribution implied by m must approach the size distribution among entrants. This distribution, denoted by G , follows from the productivity distribution J among firms attempting entry via $J(Z) = G(S[Z])$. This implies the boundary condition

$$(15) \quad \lim_{a \downarrow 0} \int_b^s m(a, x) dx = G(s) - G(b)$$

for all $s > b$. An additional boundary condition is given by the requirement that

$$(16) \quad m(a, b) = 0$$

for all $a > 0$. This condition arises from the fact that firms exit at b while none enter starting with a size below b .

LEMMA 1. The solution to (14) subject to the boundary conditions (15) and (16) is

$$m(a, s) = \int_b^\infty e^{-\eta a} \psi(a, s|x) dG(x)$$

for all $a > 0$ and all $s > b$ where

$$\psi(a, s|x) = \frac{1}{\sigma \sqrt{a}} \left[\phi \left(\frac{s - x - \mu a}{\sigma \sqrt{a}} \right) - e^{-\mu(x-b)/(\sigma^2/2)} \phi \left(\frac{s + x - 2b - \mu a}{\sigma \sqrt{a}} \right) \right],$$

and where ϕ is the standard normal density.

11. See Feller [1971] and Dixit and Pindyck [1994] for applications to industry equilibrium.

This solution can be found in Harrison [1985, p. 46] for the case of no population growth and G equal to a point mass. The two terms that define $e^{-\eta a} \psi(a, s|x)$ both satisfy (14). For small values of a , the first term approximates a normal probability density that puts almost all probability close to $s = x$. The second term converges to zero as a goes to zero, since $s + x > 2b$. This implies the boundary condition (15). The fact that $\psi(a, b|x) = 0$ for $a > 0$ implies (16). Together with $\eta \geq 0$, Assumption 4 suffices to ensure that $e^{-\eta a} \psi(a, s|x)$ can be integrated over all $a > 0$ and $s > b$ so that the overall measure of firms is finite. The following remark will be used to further characterize m .

REMARK. The roots of the characteristic polynomial $-\eta + \mu z + z^2 \sigma^2/2$ of (14) are α and $-\alpha_*$, where

(17)

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

Since $\eta \geq 0$, both roots are real, and Assumption 4 is equivalent to $\alpha > 1$. If $\eta = 0$, then α simplifies to $\alpha = -\mu/(\sigma^2/2)$. The root α_* is nonnegative and positive if and only if $\eta > 0$. If $\mu < 0$, then α_*/η converges to $1/(-\mu)$ as η goes to zero.

Observe that $m(a, s)$ reduces to $e^{-\eta a} \psi(a, s|x)$ if G is replaced by a distribution concentrated at x . This means that $e^{-\eta a} \psi(a, s|x)$ is the density of firm age and size among all firms with the same initial size x . Let $\pi(a, s|x)$ denote the associated probability density. Integrating $e^{-\eta a} \psi(a, s|x)$ to obtain the normalizing constant yields

$$\pi(a, s|x) = \left(\frac{1 - e^{-\alpha_*(x-b)}}{\eta}\right)^{-1} e^{-\eta a} \psi(a, s|x).$$

Combining this with the solution for $m(a, s)$ gives

$$(18) \quad m(a, s) = \int_b^\infty \pi(a, s|x) \left(\frac{1 - e^{-\alpha_*(x-b)}}{\eta}\right) dG(x).$$

Thus, $m(a, s)$ is a weighted sum of the densities $\pi(a, s|x) dG(x)$, with weights that are increasing in the distance of the initial size x from the exit barrier b . In the special case of $\eta = 0$, these weights reduce to $(x - b)/(-\mu)$, which is the expected life span of

a new firm entering with size x . Relatively large entering firms stay around longer and appear more often in the population than suggested by the size distribution of entrants.

III.A. The Age Distribution

If heterogeneity among entrants is small relative to heterogeneity in the overall population, then the age distribution will look much like the one obtained by conditioning on a typical $x > b$. Integrating $\pi(a,s|x)$ over s gives the age density among firms with the same size at entry. The result is

$$\pi(a|x) = \left(\frac{1 - e^{-\alpha_*(x-b)}}{\eta} \right)^{-1} e^{-\eta a} \Lambda(a|x)$$

where

(19)

$$\Lambda(a|x) = \Phi\left(\frac{x - b + \mu a}{\sigma \sqrt{a}}\right) - e^{-\mu(x-b)/(\sigma^2/2)} \Phi\left(\frac{\mu a - (x-b)}{\sigma \sqrt{a}}\right),$$

and where Φ is the standard normal distribution function. The function $\Lambda(\cdot|x)$ is the survivor function of a cohort of firms with the same initial size x .¹² If there is no population growth, then $\pi(a|x)$ is simply the survivor function scaled by the average life span of a firm. Note that $\Lambda(a|x)$ converges to $\max\{0, 1 - e^{-\mu(x-b)/(\sigma^2/2)}\}$ when age grows without bound. If $\mu \leq 0$, then all firms with a given entry size eventually exit, while a positive fraction survives forever if $\mu > 0$.

III.B. The Size Distribution

The firm size density is a weighted average of the densities $\pi(s|x)$ of size conditional on initial size. For any $x > b$, integrating $\pi(a,s|x)$ over all ages gives

(20)

$$\pi(s|x) = \left(\frac{e^{\alpha_*(x-b)} - 1}{\alpha_*} \frac{e^{\alpha(s-b)}}{\alpha} \right)^{-1} \min\left\{ \frac{e^{[\alpha+\alpha_*(s-b)]} - 1}{\alpha + \alpha_*}, \frac{e^{[\alpha+\alpha_*(x-b)]} - 1}{\alpha + \alpha_*} \right\}$$

for all $s \geq b$. This is a well-defined density for any $\alpha > 0$ and α_*

12. The size density at age a of firms of the same cohort and initial size x then satisfies (14) with η set equal to zero, and the age-zero boundary condition is a point mass at x . From this the result follows.

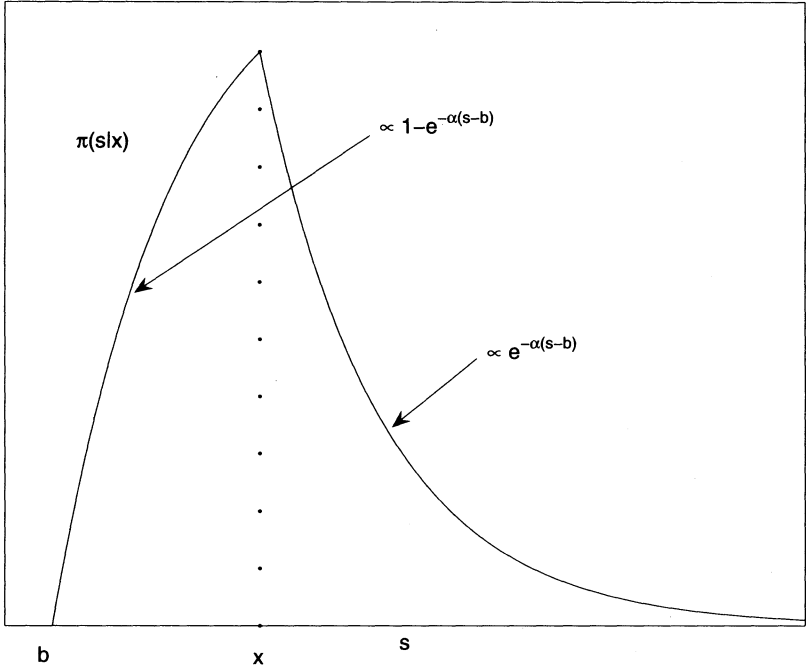


FIGURE II
Size Density Conditional on Initial Size

≥ 0 . The mean of firm size, when size is measured by e^s , is finite if and only if $\alpha > 1$. As noted earlier, this is guaranteed by Assumption 4. An example of $\pi(s|x)$ is given in Figure II. The kink at $s = x$ is a result of the entry that takes place at x . Conditional on $s \geq x$, the density of e^s implied by (20) is a Pareto density with tail probabilities of the form $e^{-\alpha(s-x)}$. The parameter α is the tail index of the conditional size distribution $\pi(s|x)$.¹³

If all new firms enter with the same initial productivity, then G is a point mass at some initial size x . In that case, (18) implies

13. Suppose population growth rates are zero. Consider the limiting distribution obtained by letting x go to b . This turns the profitability process of a dynasty of firms into a Brownian motion with a negative drift and a reflecting barrier at b . The resulting distribution for e^s is a Pareto distribution on $e^s \geq e^b$ with mean $e^b \alpha / (\alpha - 1)$. In Gabaix [1999], e^s is the size of a city relative to the average city size. This must have mean 1, and so $\alpha = 1 / (1 - e^b)$. The explanation given in Gabaix [1999] for Zipf's law for relative city sizes is that b must be very small.

that $\pi(s|x)$ is the firm size density. This density closely matches the data presented in Figure I if $x - b$ is small and $\alpha \approx 1.06$. More generally, suppose that G is a distribution with few firms that are much larger than the exit barrier. Then the size marginal $m(s)$ will inherit the exponentially declining tail common to all $\pi(s|x)$ over most of the support (b, ∞) . The deviations from linearity seen in Figure I occur for small firms: there are fewer of them than would be the case if the size distribution was Pareto. Since $\pi(s|x)$ is upward sloping on the interval (b, x) , this is exactly what is predicted when G tends to have most of its mass close to the exit barrier.

To emphasize the importance of randomness in shaping the firm size distribution, it is instructive to consider what happens as the variance of productivity shocks goes to zero. For simplicity, suppose that $\eta = 0$. Assumption 4 then requires $\mu < 0$, and at $\sigma^2 = 0$, one obtains $\xi = (r - \kappa)/|\mu|$ and $b = 0$. Firms exit immediately when they no longer break even. There is no option value that would justify continuing to operate a loss-making firm. An entering firm starts with size x , and this size will then decline linearly to 0, at which point the firm exits. One can verify that the size distribution converges to a uniform distribution on $(0, x)$ as σ^2 goes to 0. For very small σ^2 , most firms are less profitable and smaller than the most recent entrant. This is in sharp contrast to what is found in the data [Dunne, Roberts, and Samuelson 1988, 1989; Caves 1998]. The randomness in productivity growth generates a selection mechanism by which the typical firm can be much larger and productive than recent entrants.

IV. THE BALANCED GROWTH PATH

Per capita consumption and wages grow at the rate κ given by (6). The resulting interest rate is $r = \rho + \gamma\kappa$, and together with κ this pins down the value function $V(s)$. The zero-profit condition then determines $(C/w)/w^{\beta/(1-\beta)}$ and thereby the function $S[Z]$ that relates size to productivity. The resulting size distribution of firms was described in the preceding section.

It remains to determine the levels of per capita consumption and wages as well as the rate I at which firms attempt to enter. These variables are implied by market clearing conditions in the goods and labor markets. Let $L_E e^{\eta t}$, $L_F e^{\eta t}$, and $L e^{\eta t}$ denote the amounts of labor assigned to, respectively, setting up new firms,

fixed costs to operate existing firms, and production. It follows from the firm decision rules (7)–(9) that

$$(21) \quad [L_E \quad L_F \quad L] \\ = \left[\lambda_E \quad \lambda_F \int_b^\infty m(s) ds \quad \lambda_F \left(\frac{\beta}{1-\beta} \right) \int_b^\infty e^s m(s) ds \right] I.$$

Together with the labor market clearing condition $L_E + L_F + L = H$, this determines the attempted entry rate I . Aggregate output is the sum of firm revenues. The decision rules (7)–(9) imply that aggregate output $Ye^{(\kappa+\eta)t}$ satisfies

$$(22) \quad \frac{Y}{w} = \frac{\lambda_F I}{1-\beta} \int_b^\infty e^s m(s) ds.$$

In combination with the goods market clearing condition $C = Y$, this determines the ratio C/w . Since $(C/w)/w^{\beta/(1-\beta)}$ is determined by the zero-profit condition, this pins down C and w . This leads to the first part of the following proposition.

PROPOSITION 1. If Assumptions 1–4 hold, then there exists a balanced growth path. A proportional reduction in the entry and fixed cost parameters (λ_E, λ_F) raises the level of output with an elasticity $(1-\beta)/\beta$.

At $t = 0$, the distribution of productivities available to potential entrants is $J(Z)$. At that same time, there will be some measure of incumbent firms with given levels of productivity. The balanced growth path of Proposition 1 will be an equilibrium if at $t = 0$ the density of productivity among incumbent firms is $m(S[Z])/|DS[Z]|$. What happens for different initial conditions is not known.

To see the second part of Proposition 1, observe that a proportional reduction in (λ_E, λ_F) does not affect the zero-profit condition. The function $S[Z]$ and the size density $m(s)$ therefore do not change. It follows from (21) and the labor market clearing condition that I increases in such a way that $(\lambda_E, \lambda_F)I$ remains constant. Together with (22) and $C = Y$ this implies that C/w remains unchanged. Since $S[Z]$ is proportional to $(1/\lambda_F)(C/w)/w^{\beta/(1-\beta)}$, it follows that $1/w$ must increase with an elasticity $(1-\beta)/\beta$. This is also the effect on consumption. Lower setup and fixed costs imply a larger number of firms. Since firms are identified

with distinct differentiated goods, this means a larger number of goods. The elasticity $(1 - \beta)/\beta$ measures the increase in composite consumption arising from this increase in variety.

Note that (21) and (22) depend on $(\lambda_E, \lambda_F)/H$ when labor and output are expressed in per capita terms. Also, the function $S[Z]$ can be written in terms of C/H and λ_F/H . Thus, an increase in the size of the population is equivalent to a proportional reduction in the setup and fixed costs. The resulting elasticity $(1 - \beta)/\beta$ of per capita consumption with respect to H corresponds to the one obtained for the growth rate κ in (6). The benefits of lower setup and fixed costs and larger population sizes derived here replicate those obtained for a static economy by Krugman [1979].

V. IMPERFECT IMITATION—ENDOGENIZING THE TAIL INDEX

The equilibrium constructed in Proposition 1 relies on the assumption that the tail index α of the conditional size distribution $\pi(s|x)$ is greater than one. The data in Figure I suggest that α should be close to one. The parameter α is a function of the population growth rate η , the curvature parameter β of the utility function, and the technology parameters $[\theta_E, \theta_I, \sigma_Z]$. So far, these parameters have been taken as exogenous, and the model can explain Figure I only if these parameters happen to be of just the right magnitude to imply $\alpha \approx 1.06$. This section makes the trend parameter θ_E of the distribution of entry productivity endogenous and gives conditions under which the resulting equilibrium tail index will be only slightly above one.

By paying fixed costs, incumbent firms can continue to produce and generate stochastic productivity improvements. The productivity of surviving firms will tend to grow forever as long as the within-firm growth rate of productivity θ_I is not too small. If new firms had to start from the same level of productivity as existing firms entered with in the past, then the value of entry would eventually become too small to justify the cost of entry. The high productivity of successful survivors would drive up wages beyond the level at which it would be profitable for new firms to enter. The size distribution of firms would be nonstationary.

To avoid this outcome, some mechanism is needed that allows potential entrants to benefit from the productivity improvements obtained by incumbents. The mechanism proposed here is imitation. Suppose potential entrants can pay the entry cost λ_E to select a random incumbent firm and then adopt a scaled-down

version of its technology. More precisely, if the randomly selected firm at time t has a productivity $Xe^{\theta_E t}$, then the potential entrant obtains a technology capable of producing a new good with productivity $Ze^{\theta_E t} = Xe^{\theta_E t - \delta(1-\beta)/\beta}$. The parameter δ measures how much the productivity of the potential entrant will be below that of the incumbent. It is taken to be nonnegative so that imitation is imperfect. Imitation is difficult if δ is large. The implied initial size of the potential entrant is $S[Z] = S[X] - \delta$, and the entry attempt is successful if this exceeds b .

In this mechanism, random sampling and imitation tie the expected size and profitability of a potential entrant to the average size and profitability of incumbents. This sets up strong incentives for entry when the average incumbent becomes large and profitable. The result is a stationary size distribution with a well defined and finite average firm size.¹⁴

V.A. The Stationary Size Distribution

Suppose the cross-sectional distribution of productivity is stationary when productivity is de-trended by some growth rate θ_E , to be determined in Section V.B. Suppose further that the resulting size distribution has a probability density $f(s)$. The mechanism by which potential entrants obtain a new technology implies a size density for firms attempting entry equal to $DG(x) = f(x + \delta)$, $x > b - \delta$. Integrating (14) over all ages and using the boundary condition (15) gives, for all $s > b$,

$$(23) \quad \eta f(s) = -\mu Df(s) + \frac{1}{2} \sigma^2 D^2 f(s) + \epsilon_A f(s + \delta),$$

where $\epsilon_A = 1/\int_b^\infty m(x) dx$ is the rate at which new firms attempt to enter, as a fraction of the number of existing firms. Note that ϵ_A must exceed η if the number of firms is to grow at a rate η .

LEMMA 2. Suppose $\mu < \delta\eta$, and let $\epsilon_A > \eta$ be the unique entry rate for which the characteristic equation $\eta = \mu z + \sigma^2 z^2/2 + \epsilon_A e^{-\delta x}$ has only one solution. This solution is given by $z = \zeta$, where

14. In Eaton and Eckstein [1997], knowledge spillovers across existing cities provide the mechanism by which the size distribution of cities is prevented from spreading out. Jovanovic and MacDonald [1994] and Eeckhout and Jovanovic [2002] allow all firms to copy, imperfectly, from the whole population of firms. Here, the spillover is only from incumbents to potential entrants. Incumbents are locked into their idiosyncratic productivity processes and are not assumed to be able to imitate the successes of other incumbent firms.

$$(24) \quad \zeta = -\left(\frac{\mu}{\sigma^2} + \frac{1}{\delta}\right) + \sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + \frac{1}{\delta^2} + \frac{\eta}{\sigma^2/2}}.$$

Then the stationary density that solves (23) together with the boundary condition $f(b) = 0$ is the gamma density

$$(25) \quad f(s) = \zeta^2(s - b)e^{-\zeta(s-b)}.$$

For $\delta = 0$, (24) is understood to represent the limiting value $\zeta = -\mu/\sigma^2$. One can derive (24) by minimizing the right-hand side of the characteristic equation. The condition $\mu < \delta\eta$ is necessary and sufficient to ensure that $\zeta > 0$. The tail probabilities of $f(s)$ behave like $e^{-\zeta s}$ for large s , and so ζ does indeed represent the tail index of the size distribution. For large δ entrants tend to be small and the tail index ζ is essentially the same as the tail index α of the conditional size distribution $\pi(s|x)$. The right-hand side of (24) is decreasing in μ and thus increasing in the growth rate θ_E . The higher the average growth rate θ_E of productivity in the population relative to the drift θ_I of surviving incumbents, the more aggregate productivity growth must be due to selection, and this implies a size distribution with a thinner tail. The mean of e^s implied by $f(s)$ is finite if and only if $\zeta > 1$.

Lemma 2 defines a particular entry rate ε_A and solves (23). For any other $\varepsilon_A > \eta$, the differential equation (23) is solved, subject to the boundary condition $f(b) = 0$, by $zz_*[e^{-z(s-b)} - e^{-z_*(s-b)}]/(z_* - z)$, where $z \in \mathbb{C}$ and $z_* \in \mathbb{C}$ solve the characteristic equation defined in Lemma 2. Proper densities arise when ε_A is such that z and z_* are real. To motivate focusing on the $z = z_* = \zeta$ solution shown in (24) and (25), consider a new “industry” of many firms that all start out with the same initial size $x > b$. Suppose that over time new firms attempt to enter this industry at some rate ε_A by imitating incumbents in the industry, as described above. Let $n(a,s)$ be the size density of firms in this industry at age a . Then $n(a,s)$ satisfies $D_a n(a,s) = -\mu D_s n(a,s) + \sigma^2 D_{ss} n(a,s)/2 + \varepsilon_A n(a,s + \delta)$ and $n(a,b) = 0$. Consider the special case $\delta = 0$ and take the initial measure of firms to be one. The solution for $n(a,s)$ is then given by $n(a,s) = e^{\varepsilon_A a} \psi(a,s|x)$. Normalizing this solution by the number of firms yields a distribution that converges to the gamma distribution (24) and (25) as the industry ages. This is also true when entry rates vary over time and when firms at the initial date differ in size, as long as the initial size distribution has a support that is

compact and contained in (b, ∞) . Thus, compactly supported initial size distributions converge to (24) and (25) and not to the other solutions of (23). Given the limiting distribution (24) and (25) generated by the process of selection and imitation, the entry rate ϵ_A defined in Lemma 2 is simply the rate required to make the number of firms grow at a rate η .¹⁵

V.B. The Balanced Growth Path and Zipf's Law

The size density $f(s)$ constructed in Lemma 2 is a function of the assumed productivity growth rate θ_E through its dependence on the drift parameter μ . The value function $V(s)$ is also a function of θ_E , via μ , as well as via the equilibrium interest rate r and the growth rate κ of per capita consumption and wages. Taken together, this means that the expected profits from entry are a function of θ_E . The only values of θ_E that are consistent with balanced growth are those for which these profits are zero:

$$(26) \quad \lambda_E = \lambda_F \int_b^\infty V(x) f(x + \delta) dx.$$

Together with (24) and (25), this zero-profit condition determines θ_E and $f(s)$. Taking $DG(x) = f(x + \delta)$ in Lemma 1 gives the density $m(s)$ of firms per entry attempt, and inserting $f(s)$ into the differential equation (23) yields the equilibrium attempted entry rate ϵ_A . To complete the construction of a balanced growth path, recall that the relation between firm size s and productivity $Ze^{\theta_E t}$ is determined by $s = S[Z]$. From the definition (9), $e^{S[Z]}$ is proportional to $(C/w)/w^{\beta/(1-\beta)}$. The location of the productivity density $f(S[Z])|DS(Z)|$ is therefore determined by the log of $(C/w)/w^{\beta/(1-\beta)}$. On a balanced growth path, the density $f(S[Z])|DS(Z)|$ must correspond to the density of productivity among incumbent firms at the initial date, which is an initial condition for the economy. Assuming that the productivity distribution at the initial date is consistent with balanced growth, this

15. With perfect imitation, the density $n(a, s)$ has a spectral representation [Karlin and Taylor 1981, p. 393] consisting of eigenfunctions of the right-hand side of (23). The underlying reason for the convergence to (24) and (25) is that this density is the eigenfunction associated with the supremum of the eigenvalues that appear in this representation. The technical appendix available at www.luttmer.org proves the convergence to (24) and (25) and this interpretation. The stability argument described here covers only the case $\delta = 0$ and does not explain why μ and b are constant parameters. A more complete analysis of stability awaits further research.

requirement determines the equilibrium value of $(C/w)/w^{\beta/(1-\beta)}$. As in the case of exogenous growth, (21) and (22) together with goods and labor market clearing conditions determine the ratio C/w and the rate I at which firms attempt to enter. Together with $(C/w)/w^{\beta/(1-\beta)}$ this yields C and w separately, and the economy will be on a balanced growth path if the number of firms at the initial date equals $I \int_b^\infty m(s) ds = I/\varepsilon_A$.

The following proposition shows that this construction works if consumers discount the future at a high enough rate. Precise conditions and a proof are in Appendix I.

PROPOSITION 2. Suppose the population growth rate η and the drift θ_T of within-firm technological progress are nonnegative. If the discount rate ρ is large enough, then there exists a balanced growth path with a size distribution defined by (24) and (25). The tail index ζ of the size distribution converges to one—Zipf's Law—as the ratio λ_E/λ_F of entry over fixed costs grows without bound.

The existence of a balanced growth path and the circumstances in which Zipf's law arises are most transparent in the special case of logarithmic utility. This case implies that $r = \rho + \kappa$, simplifying the dependence of the value of a firm on θ_E . For fixed $u = x - b$, the value $V(u + b)$ is then unambiguously decreasing in θ_E . Higher productivity growth in the population drives incumbents at a given distance from the exit barrier out of business more quickly, and this implies a low firm value. As noted earlier, a higher θ_E generates a size distribution with a thinner tail, or a higher ζ . High- ζ gamma densities (25) are stochastically dominated by low- ζ gamma densities in a first-order sense. Since $V(u + b)$ is an increasing function of u , it follows that the right-hand side of the zero-profit condition (26) is decreasing in θ_E .¹⁶ Equivalently, the expected value of entry is decreasing in the tail index ζ . It is not difficult to show that the value of entry goes to zero for very large ζ . Finally, the dominant term in the value function $V(x)$ is the firm size variable e^x , and this implies that the expected value of entry grows without bound as the tail index ζ approaches 1 from above. The right-hand side of the zero-profit condition is therefore as shown in Figure III, with a

16. Faster growth increases the exit barrier b , and this tends to shift the size distribution to the right. But, because entrants sample from the population of incumbents, what matters for the value of entry is the distribution of size relative to the exit barrier.

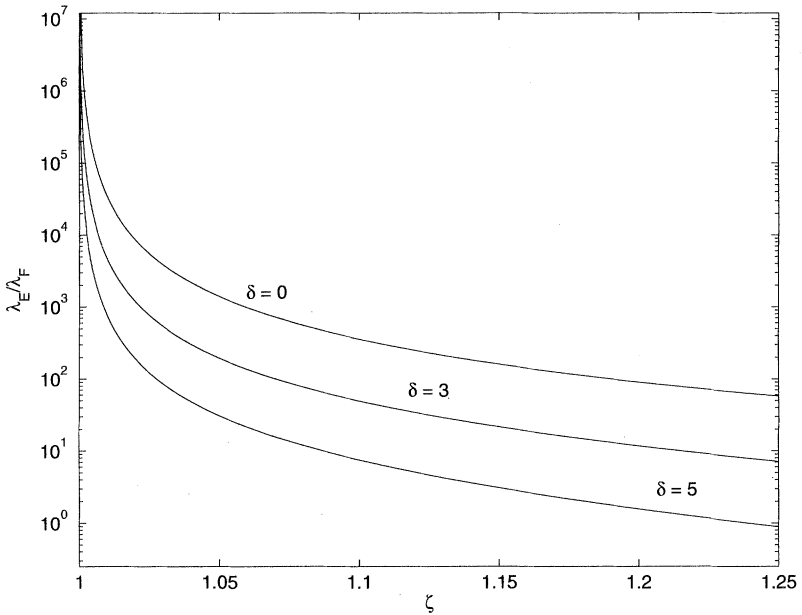


FIGURE III
Entry Costs, Fixed Costs, and the Tail Index

vertical asymptote at $\zeta = 1$ and a horizontal asymptote at 0. From this the results of Proposition 2 follow.¹⁷

If the utility function exhibits more curvature than logarithmic utility, then the value function continues to be monotone in ζ for high enough discount rates. But if $\gamma < 1$, then the discount factor $1/(r - \kappa)$ is increasing in κ and thus also in θ_E and ζ . This can outweigh the negative effect on the value function of a larger gap $\theta_E - \theta_I$ between productivity growth in the population and the drift of incumbent productivity. The value of a firm may, over some range, increase with the growth rate of productivity in the population. This can make the expected value of entry nonmonotone in θ_E and ζ . The proof given in Appendix I shows that a balanced growth path does nevertheless exist for high enough discount rates ρ .

17. The parameters for Figure III are taken from the calibration in Section VI.A, assuming $\gamma = 1$ and using an interest rate of 4 percent per annum.

V.C. Barriers to Entry and Growth

The equilibrium conditions (23) and (26) and, therefore, the growth rate θ_E are independent of the scale of the entry and fixed costs (λ_E, λ_F). As in the case of exogenous growth, lowering both costs at the same time increases the level of output with an elasticity $(1 - \beta)/\beta$. The effects of changing only barriers to entry—the entry cost λ_E or the difficulty of imitation δ —are described in the following corollary of Proposition 2.

COROLLARY. Suppose the conditions of Proposition 2 hold. The growth rate θ_E of productivity in the population is decreasing in the entry cost parameter λ_E and the imitation parameter δ when $\gamma \geq 1$ and for sufficiently large entry costs when $\gamma < 1$.

For $\gamma \geq 1$, this result follows from the fact that the value of entry, as illustrated in Figure III, is decreasing in the tail index ζ . A higher entry cost λ_E implies a higher equilibrium value of entry and, thus, a lower equilibrium value of ζ and a lower θ_E . Similarly, a larger δ implies a lower equilibrium value of ζ since the expected value of entry is lower when imitation is more difficult. Given that the right-hand side of (24) is increasing in δ and decreasing in μ , this implies a lower growth rate θ_E . For $\gamma < 1$, these conclusions continue to hold provided entry costs are high. High entry costs imply that ζ must be close to 1 and the expected value of entry can be shown to be monotone for all ζ close enough to the asymptote $\zeta = 1$.

If imitation is difficult and population growth is small, then (24) implies that $\zeta \approx -\mu/(\sigma^2/2)$. Together with the definitions (10) of μ and σ^2 , this yields a simple expression relating the equilibrium productivity growth rate θ_E and the equilibrium tail index ζ :

$$(27) \quad \theta_E \approx \theta_I + \frac{\zeta\beta}{1-\beta} \frac{\sigma_Z^2}{2}.$$

The drift of incumbent productivity is θ_I , and the second term in (27) captures the effect of selection on productivity growth in the population of firms. Lower barriers to entry imply smaller firms, and this corresponds to higher values of ζ . By (27), this means faster productivity growth in the population. Incumbent productivity drifts up at a rate θ_I in any case, but the lower barriers to entry generate more firm turnover, and this increases the effect of selection.

V.D. Firm Exit Rates by Age

The specific size distribution for entering firms implied by imitation generates a precise prediction about the dependence of firm exit rates on age. The main properties of the hazard rate are summarized in the following proposition. Explicit formulas and a sketch of the proof are given in Appendix II.

PROPOSITION 3. If $\delta = 0$, then firms exit from a given age cohort with a hazard rate that is independent of age. If $\delta > 0$, then the hazard rate $h(a)$ is strictly decreasing and satisfies

$$\lim_{a \downarrow 0} h(a) = \infty, \quad \lim_{a \rightarrow \infty} h(a) = \frac{1}{2} \left(\left[\frac{-\mu}{\sigma} \right]^+ \right)^2.$$

For given $x > b$, the hazard rate of the conditional survivor function $\Lambda(a|x)$ defined in (19) is a hump-shaped function of age and zero at age zero. Firms entering with a productivity that exceeds the exit barrier by a certain amount do not exit initially. As these firms are subjected to productivity shocks, some start to exit, and the hazard rate increases. Eventually, sufficiently many surviving firms will have moved away from the exit barrier as a result of favorable productivity shocks, and the hazard rate declines again. In contrast, firms in a cohort of imitating entrants come with initial sizes x that are arbitrarily close to the exit barrier b , and so significant exit will take place right from the start. If new entrants can perfectly copy a randomly selected incumbent, then the rate at which firms exit is not hump-shaped but constant. If imitation is imperfect, then entrants tend to be smaller than incumbents. The probability of exit decreases with size, and it takes time for firms to grow. The result is an exit rate that declines with age.¹⁸

VI. CALIBRATIONS

Growth is due to increased variety, within-firm technological progress, and selection. This section describes how the observed size distribution together with entry or exit data can be used to

18. Caves [1998] discusses the literature on firm exit rates and cites studies documenting hazard rates that decline with age. Based on monthly observations of a cohort of new firms in the Munich (Germany) area, Brüderl, Preisendörfer, and Ziegler [1992] report a hump-shaped hazard function. Nucci [1998] finds a hump-shaped hazard function for establishments that peaks around an age of one year.

infer the magnitude of these different sources of growth under the assumption that preferences are described by $\beta = 0.9$. This benchmark value implies that the differentiated goods produced by different firms are close substitutes and that variable profits are relatively small. Data on revenues and variable costs could be used to determine β . Alternatively, β could be identified from the demand curves (2) using price and quantity observations on the composite goods sold by individual firms, and instruments correlated with technology shocks but not taste shocks. A careful investigation along these lines is beyond the scope of this paper.

VI.A. *Inferring the Contribution of Selection to Growth*

The regression line through all the data points in Figure I that represent five or more employees has a slope of -1.06 , suggesting that $\zeta \approx 1.06$. A comparison of the size distributions of incumbents and entrants can be used to infer the imitation parameter δ . The statistics reported in Figure I imply that 87.7 percent of all firms with at least one employee had fewer than twenty employees in 2002. For new employer firms, this fraction was 95.0 percent.¹⁹ These two fractions together with $\zeta \approx 1.06$ imply that $\delta \approx 3$.²⁰ This estimate means that the size of an imitating entrant is less than 5 percent of the size of the incumbent being imitated.

To decompose the economy-wide rate of technological progress θ_E into a within-firm growth rate θ_I and a selection component $\theta_E - \theta_I$ requires an estimate of μ . When η is small and δ is large, the definition (24) of ζ implies $-\mu \approx \zeta\sigma^2/2$. The variance σ^2 of firm growth can be identified from the rate ε_S at which new firms succeed to enter per unit of time, relative to the total number of firms. This entry rate equals the population growth rate plus the exit rate. The rate at which firms cross the exit barrier b is given by $Df(b)\sigma^2/2$,²¹ and, therefore, $\varepsilon_S = \eta +$

19. See Table 743 of the 2006 edition of the Statistical Abstract of the United States.

20. Among entrants, the fraction of firms of size at least s is $[1 + \zeta(s - b)/(1 + \delta\zeta)]e^{-\zeta(s-b)}$. Setting δ equal to zero in this expression gives the corresponding fraction for incumbent firms. Equating these fractions to the respective empirical fractions 0.050 and 0.123 gives $s - b = 3.42$ and $\delta = 3.04$. If the employment statistics represent variable labor, then the minimum firm size is $20e^{-3.42} = 0.65$ employees.

21. The size density of an age cohort of firms satisfies (14) with η set equal to zero. Integrating this equation over size shows that the rate at which a particular age cohort shrinks over time is proportional to the slope of the cohort size density at b . Adding up over all age cohorts then gives the result.

$\zeta^2\sigma^2/2$. The U. S. Small Business Administration reports an entry rate of 11.6 percent per annum for the year 2002. Postwar U. S. population growth is about 1 percent per annum. Together, the estimates of ζ , ε_S , and η imply

$$(28) \quad -\mu \approx \frac{\varepsilon_S - \eta}{\zeta} = 0.1, \quad \sigma^2 = \frac{\varepsilon_S - \eta}{\zeta^2/2} = (0.43)^2.$$

Solving (24) exactly with $\eta = 0.01$ and $\delta = 3$ gives the slightly more negative drift estimate of $\mu = -0.12$.

Combined with the benchmark parameter $\beta = 0.9$ and the definitions of μ and σ , these estimates imply $\theta_E - \theta_I = -\mu(1 - \beta)/\beta = 0.013$, $\sigma_Z = \sigma(1 - \beta)/\beta = 0.048$. In postwar U. S. data, the growth rate of per capita GDP is a little over 2 percent. From (6), this gives rise to the decomposition

$$\kappa = \underbrace{\theta_I}_{0.006} + \underbrace{\theta_E - \theta_I}_{-\mu(1 - \beta)/\beta = 0.013} + \underbrace{\left(\frac{1 - \beta}{\beta}\right)\eta}_{0.001} = 0.02.$$

Since goods are assumed to be close substitutes and population growth is only about 1 percent per annum, the contribution to growth of increases in variety is small. In contrast, the fact that the tail index ζ is only marginally above 1 while the entry rate ε_S is as large as 11.6 percent per annum implies that $-\mu$ and σ must be large, by (28). Selection must then play an important role, even when the differentiated commodities produced by different firms are close substitutes.

VI.B. Some Empirical Caveats

Although the gamma distribution has a right tail that can match the data, it does not quite fit the empirical size distribution shown in Figure I. If employment statistics are interpreted as variable labor, then the tail index $\zeta = 1.06$ and the observed fraction of firms with no more than twenty employees imply a minimum firm size of 0.65 employees. The resulting gamma density has too few small firms, and the implied number of firms with at least a thousand employees is more than twice as large as in the data. Alternatively, the maximum likelihood estimator based on the data shown in Figure I gives $\zeta = 1.30$ and a minimum firm size of 1.22 employees. For size categories below a thousand employees, this gamma distribution matches the data extremely

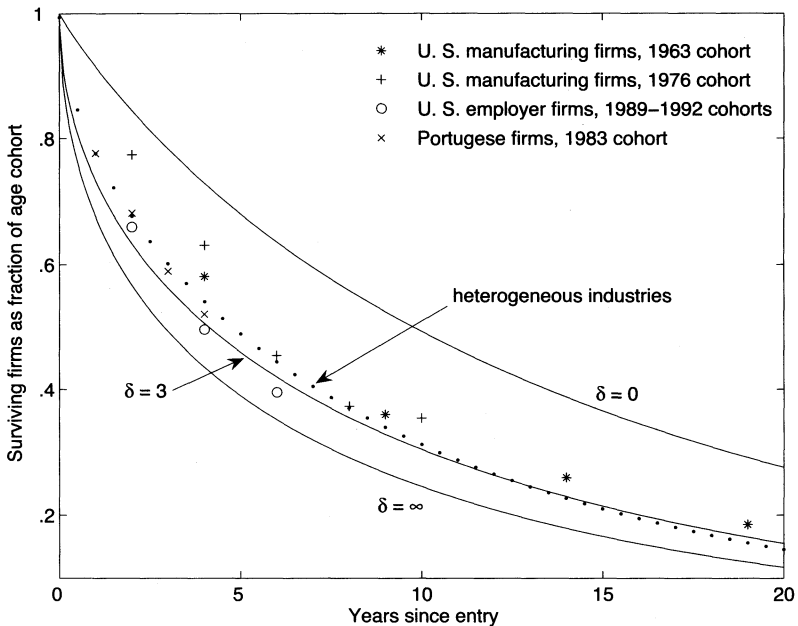


FIGURE IV
Survival Rates

well. But because the estimate of ζ is now well above 1, this distribution does not predict enough large firms.

The hazard rate implications described in Proposition 3 and Appendix II provide a further set of over-identifying restrictions. Figure IV shows the survivor function implied by $\mu = -0.12$, $\sigma = 0.43$, and $\delta = 3$, together with a number of empirical survivor functions. Included are data on the 1963 and 1976 cohorts of U. S. manufacturing firms obtained from, respectively, Dunne, Roberts, and Samuelson [1988] and Audretsch [1991]; a cohort of Portuguese manufacturing firms set up in 1983 and studied by Mata and Portugal [1994]; and a cohort of new U. S. employer firms set up in the early 1990s described in Headd [2003]. Also shown for comparison are the survivor functions that correspond to $\delta = 0$ and $\delta = \infty$, holding fixed $\mu = -0.12$ and $\sigma = 0.43$. Although there is variation in empirical survival rates that is not accounted for, the observed survival rates are in the range predicted by the model.

The estimated standard deviation of firm growth, $\sigma = 0.43$, is

surprisingly large. For small fixed costs, this standard deviation is also, approximately, the standard deviation of the stock return of a typical firm. Campbell et al. [2001] find that the annual standard deviation of the stock return is about 0.3 for the typical NYSE or NASDAQ listed firm, and most of the standard deviation is due to idiosyncratic shocks. At the cost of underpredicting the number of large firms, the maximum likelihood estimate of ζ provides a partial remedy. Given $\zeta = 1.30$, the empirical fractions of incumbent and entering firms with fewer than twenty employees imply $\delta = 2.5$, and the resulting standard deviation of firm growth shrinks to $\sigma = 0.35$. This is noticeably closer to the stock market proxy of 0.3. But leverage considerations suggest that even this proxy is only an upper bound on the standard deviation of firm growth rates. An alternative remedy is to allow for random exit by firms that are not at the exit barrier b as in Luttmer [2004]. Observed entry rates are then consistent with lower exit rates at the exit barrier, and this implies a lower variance of firm growth rates. Random exit would also imply a smaller role for selection.²²

VI.C. *Heterogeneity Across Industries*

In the economy described so far, all firms face the same demand curves, and all experience changes in demand and productivity described by the same drift and diffusion parameters. No doubt, the degree to which the differentiated commodities produced in an industry are substitutable differs across industries as do the typical rates of technological progress. Nor are entry and fixed costs or the difficulty of imitation likely to be the same across industries. It is therefore perhaps not surprising that the gamma density implied by a one-industry economy does not quite match the data in Figure I. This section shows that even a limited amount of heterogeneity across industries can be used to produce the remarkable fit shown in Figure I.²³

Consider an economy with N different goods, each of which is a composite of a continuum of differentiated commodities. Industries are identified with different composite goods. As before,

22. Evidence presented in Cabral and Mata [2003] suggests that up to 1991 exit from the 1984 cohort of new Portuguese manufacturing firms was unrelated to size. Substantial heterogeneity in fixed costs could give rise to this.

23. Luttmer [2004] allows for additional sources of within-industry heterogeneity by incorporating within-industry variation in fixed and entry costs as well as in technologies used to combine physical capital and labor to produce differentiated goods.

different firms in an industry produce distinct differentiated commodities. Consumption is given by the Cobb–Douglas aggregate $C_t = \prod_{n=1}^N C_{n,t}^{\nu_n}$, where $C_{n,t}$ satisfies (1) with β replaced by an industry-specific curvature parameter β_n . The share parameters ν_n are between zero and one and add up to one. Idiosyncratic firm productivity in industry n is assumed to follow (5), with $[\theta_E, \theta_I, \sigma_Z]$ replaced by $[\theta_{E,n}, \theta_{I,n}, \sigma_{Z,n}]$.

Along any balanced growth path, aggregate consumption of the composite good produced by industry n will be $C_{n,t} = C_n e^{(\kappa_n + \eta)t}$, where κ_n is defined in terms of $\theta_{E,n}$ and β_n as in (6). Aggregate consumption equals $C_t = C e^{(\kappa + \eta)t}$, and the growth rate κ of per capita consumption is simply $\sum_{n=1}^N \nu_n \kappa_n$, the average of the industry growth rates weighted by expenditure shares. Wages also grow at this rate. The price index for aggregate consumption is $P_t = \prod_{n=1}^N (P_{n,t}/\nu_n)^{\nu_n}$, where $P_{n,t}$ is the price index for the composite good of industry n , defined as in (3). The relative prices $P_{n,t}/P_t$ must be given by $(P_n/P) e^{(\kappa - \kappa_n)t}$, since expenditure shares are constant. Let $\lambda_{F,n}$ be the fixed cost required to continue a firm in industry n . A calculation along the lines of (7)–(9) implies that the relation between productivity and size in industry n is given by

$$e^{S_n[Z]} = \frac{\nu_n(1 - \beta_n)}{\lambda_{F,n}} \left(\frac{\beta_n Z P_n / P}{w} \right)^{\beta_n / (1 - \beta_n)} \frac{C}{w},$$

where $P = \prod_{n=1}^N (P_n/\nu_n)^{\nu_n}$. The gross revenues at time t of a firm in industry n with a productivity $Z e^{\theta_{E,n}t}$ are equal to $\lambda_{F,n} e^{S_n[Z]}$ units of labor. The (logarithmic) size of such a firm follows a Brownian motion with drift μ_n and diffusion coefficient σ_n defined as in (10), using the industry-specific parameters β_n and $[\theta_{E,n}, \theta_{I,n}, \sigma_{Z,n}]$. Firms choose to follow the same stopping rule as before, exiting when size falls below an industry-specific barrier b_n defined as in (12). The size distributions in all industries are therefore of the form derived in Section III.

Suppose firms can choose which industry to enter, and then, at a cost of $\lambda_{E,n}$ units of labor, attempt to imitate incumbents in that industry along the lines of Section V. The extent to which entrants lag behind incumbents in industry n is measured by δ_n . Potential entrants can direct their entry attempts to a specific industry, but imitation of firms in the chosen industry is imperfect, as before.

This setup leads to equilibrium conditions for the industry growth rate $\theta_{E,n}$ and size density f_n that are exactly analogous to

(23)–(26). The value functions V_n appearing in equilibrium conditions analogous to (26) depend on μ_n and the difference $r - \kappa$ between the interest rate and the aggregate growth rate κ . Since κ depends on an expenditure-weighted average of the industry growth rates $\theta_{E,n}$, this gives a system of N equilibrium conditions in N unknown growth rates $\theta_{E,n}$. For general γ , the analysis of this system is more complicated than the analysis that led to Proposition 2. But logarithmic utility implies $r - \kappa = \rho$, and then the equations uncouple: the zero-profit condition for industry n only depends on the growth rate $\theta_{E,n}$ of industry productivity and the size density f_n . As a result, the proof of Proposition 2 applies. In particular, industries with high ratios $\lambda_{E,n}/\lambda_{F,n}$ or large δ_n will have tail indices ζ_n close to 1, and, ceteris paribus, growth rates $\theta_{E,n}$ that are not far above $\theta_{I,n}$.

The overall size density will be a weighted average of the industry size densities f_n . The log of variable labor l used by a firm of size s in industry n is determined by $e^l = e^s \lambda_{F,n} \beta_n / (1 - \beta_n)$. The economy-wide density of log variable labor is therefore

$$\sum_{n=1}^N q_n f_n \left(l - \ln \left(\frac{\lambda_{F,n} \beta_n}{1 - \beta_n} \right) \right)$$

for weights q_n that add up to one. These weights are proportional to the numbers of firms in each industry. The number of firms in industry n times the average revenues in that industry should equal the value of aggregate consumption of the composite good produced in the industry, or v_n times the value of aggregate consumption. It follows that the number of firms in industry n is proportional to

$$q_n \propto v_n \left(\frac{\lambda_{F,n}}{1 - \beta_n} \int_{b_n}^{\infty} e^s f_n(s) ds \right)^{-1}$$

In other words, the number of firms in an industry is proportional to the expenditure share of that industry and inversely proportional to average gross revenues in the industry.

The curve shown in Figure I represents the size distribution of an economy with $N = 20$ and imitation parameters $\delta_n = n/4$. The imitation parameter in the most difficult industry to enter is $\delta_N = 5$, and, thus, $\delta_N(1 - \beta_N)/\beta_N = 5/9$. New firms in this industry are only about 57 percent as productive as the incum-

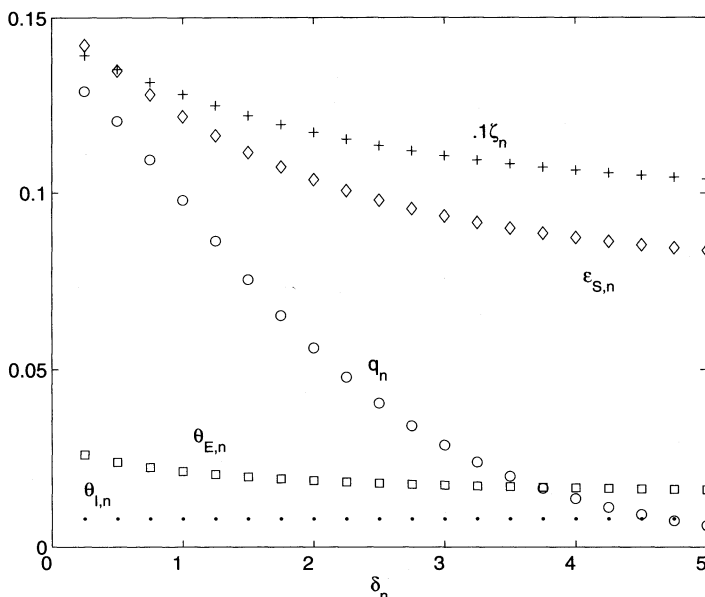


FIGURE V
Heterogeneous Industries

bents they try to imitate, and their size is less than 1 percent of the size of these incumbents. All industries have preference and technology shocks parameterized by the same $[\theta_{I,n}, \sigma_{Z,n}]$ and entry and fixed costs given by the same $\lambda_{E,n}$ and $\lambda_{F,n}$. Population growth is $\eta = 0.01$, utility is logarithmic, $\rho = 0.02$, and $\beta_n = 0.9$ as before. The values of the common $\lambda_{E,n}/(\lambda_{F,n}/\rho)$ and $\sigma_{Z,n}$ are chosen to ensure a tail index of 1.04 and an economy-wide entry rate of 11.6 percent per annum. This yields $\lambda_{E,n}/(\lambda_{F,n}/\rho) = 0.81$ and $\sigma_{Z,n} = 0.041$. The implied standard deviation of firm growth is 0.37, down somewhat from its puzzlingly high value of 0.43 in the one-industry economy.

Figure V shows the implied industry-specific tail indices ζ_n , entry rate $\epsilon_{S,n}$, and productivity growth rates $\theta_{E,n}$, as well as the fraction of firms q_n in industry n . As expected, industries in which it is easier to imitate have more entry, more rapid productivity growth through selection, and a size distribution with a thinner tail. The tail index of the overall distribution is determined by $\zeta_N = 1.04$, even though the fraction of firms in industry N is less than 1 percent. The entry rate is highest in industry 1.

Selection contributes 1.82 percent to an output growth rate of 2.72 percent in this industry, while the corresponding numbers are only 0.82 and 1.73 percent in industry N . The average contribution of selection to growth across industries is 1.09 percent, essentially the same as in the one-industry economy as is the aggregate survivor function shown in Figure IV.

The only heterogeneity across industries assumed in Figures I and V is in the imitation parameter δ_n . Because of this, larger firms tend to be in industries with low productivity growth. If, instead, industries only differ in terms of the standard deviation $\sigma_{z,n}$ of productivity shocks, then large firms would tend to be in the high- $\sigma_{z,n}$ industries where selection produces high productivity growth. Other possible sources of variation are the drift of incumbent productivity growth, within-industry substitutability of the differentiated commodities, and fixed and entry costs. Rossi-Hansberg and Wright [2004] document how size distributions vary across industries. Further research is needed to see if and how this variation can be accounted for using the model economy described here, augmented with the additional sources of within-industry heterogeneity described in Luttmer [2004].

VII. CONCLUDING REMARKS

If new entrants can imitate incumbents, then growth is rapid when barriers to entry are low. The engine of growth is experimentation by firms combined with selection. Lucky firms receive another draw, and unlucky ones exit and are replaced by more productive firms. Firms are experiments that can be cut short and replaced by new ones when they do not perform well. Reducing the cost of entry speeds up the rate of economy-wide experimentation and raises the growth rate of the economy. The resulting size distribution is stationary because potential entrants can learn from successes achieved by incumbents. It has a very thick tail when entry is difficult, nevertheless.

This model is consistent with three first-order features of the data. The economy grows at a steady rate. Firm exit rates are high for young firms and low for firms that have survived for some time. The predicted size distribution of firms closely approximates Zipf's law if entry is difficult. This tends to be true even if entry is easy in some industries.

The closed-form solutions derived in this paper rely heavily on the absence of aggregate uncertainty and on the use of steady

states. This precludes an analytical treatment of transitions and of the possible role of selection and imitation in speeding up transitions. An important abstraction also is that every firm is identified with a technology to produce a single differentiated good. In contrast, the empirical definition of a firm is based on the legal criterion of ownership. Building models of firm dynamics in which the definition of a firm corresponds more closely to the empirical definition remains an important task for further research.

In this paper, the variance of firm growth rates over small intervals of time is the same for firms of all sizes. Many studies have found larger variances for small firms than for large firms. One possible explanation for this phenomenon is the presence of unobservable fixed effects about which young firms learn, as proposed by Jovanovic [1982].²⁴ This can be combined with the permanent shocks emphasized in this paper, although the resulting hybrid model does not appear to be analytically tractable. Pakes and Ericson [1998] derive observable implications for such a hybrid model and present evidence that the importance of learning varies across industries.

APPENDIX I

Proof of Proposition 2. The following assumptions are maintained throughout:

$$(29) \quad \eta \geq 0, \theta_I \geq 0, \sigma^2 > 0, \delta \geq 0.$$

A. Existence

It is convenient to solve for the equilibrium value of μ . The growth rates κ and θ_E and the parameter ξ then follow from (6), (10), and (12). The interest rate is given by $r = \rho + \gamma\kappa$. The present value of the aggregate labor endowment must be finite in any equilibrium. Along a balanced growth path, this requires that $r > \kappa + \eta$.

LEMMA A1. If $r > \kappa + \eta$ then $\zeta > 1$ implies $r - \kappa > \mu + \sigma^2/2$.

This lemma ensures that the value function $V(s)$ given in (11) is well defined whenever the present value of the aggregate labor

24. The fact that the variance of firm growth rates is decreasing in size is emphasized in Sutton [2002] and Klette and Kortum [2004], who provide alternative interpretations.

endowment is finite and $\zeta > 1$. Under these conditions, the zero-profit condition (26) can be written as

$$(30) \quad \frac{\lambda_E}{\lambda_F} = \frac{\xi e^{-\zeta\delta}}{r - \kappa} \times \frac{(\zeta - 1)\zeta(\zeta + \xi)\delta + \zeta(\zeta + \xi) + (\zeta - 1)(\zeta + \xi) + (\zeta - 1)\zeta}{(\zeta - 1)^2(\zeta + \xi)^2}.$$

If $\xi > 0$, $r > \kappa$, and $\zeta > 1$, then the right-hand side of (30) is increasing in ξ and decreasing in $r - \kappa$ and ζ .

The definition (24) implies that ζ is strictly decreasing in μ , with a horizontal asymptote at $-1/\delta$ for large μ . Furthermore, ζ can be made arbitrarily large by taking μ small enough. The condition $\zeta > 1$ corresponds to $\mu < \mu^*$ where

$$\mu^* = \frac{\delta\eta - (1 + \delta/2)\sigma^2}{1 + \delta}.$$

The parameter ξ defined in (12) depends on μ , both directly and via

$$r - \kappa = \rho + (\gamma - 1) \left[\theta_I + \left(\frac{1 - \beta}{\beta} \right) (\eta - \mu) \right].$$

The overall dependence of ξ on μ is characterized in the following lemma.

LEMMA A2. If $r > \kappa$, then ξ is strictly increasing in μ for all $\gamma \in (0, 1]$, and for all $\gamma \in (1, \infty)$ such that

$$(31) \quad \rho > (1 - \gamma) \left[\theta_I + \left(\frac{1 - \beta}{\beta} \right) \eta \right] + \frac{1}{2} (1 - \gamma)^2 \sigma_Z^2.$$

Existence of an equilibrium will now be shown separately for $\gamma = 1$, $\gamma > 1$, and $\gamma < 1$.

Suppose $\gamma = 1$. This implies $r - \kappa = \rho$, and a necessary condition for a balanced growth path to exist is $\rho > \eta$. This condition is also sufficient. To see this, first recall from (12) and (24) that ζ is decreasing and ξ is increasing in μ . Furthermore, ζ grows without bound and ξ goes to zero as μ goes to $-\infty$. It follows that the right-hand side of (30) is an increasing function of μ , with a vertical asymptote at μ^* and a horizontal asymptote at 0.

Next suppose $\gamma > 1$. Note that $r - \kappa$ is decreasing in μ . Assume that (31) holds. Then the right-hand side of (30) is increasing in μ as long as $r > \kappa$. As μ goes to $-\infty$, $r - \kappa$ will become

large, and the right-hand side of (30) goes to zero. As μ approaches μ^* , the right-hand side will increase without bound as long as $r > \kappa$. It follows that the zero-profit condition will have a unique solution if $\mu < \mu^*$ also guarantees $r > \kappa + \eta$. This is the case if

$$(32) \quad \rho > \eta + (1 - \gamma) \left(\theta_I + \frac{1 - \beta}{\beta} \frac{\eta + (1 + \delta/2)\sigma^2}{1 + \delta} \right).$$

There is an equilibrium if ρ is large enough to satisfy both (31) and (32).

Finally, suppose $\gamma \in (0,1)$. Now there is a lower bound μ_* so that $r > \kappa + \eta$ if and only if $\mu > \mu_*$. A necessary condition for the existence of an equilibrium is therefore $\mu_* < \mu^*$. This is guaranteed if (32) holds. As μ approaches μ^* from below, ζ approaches 1 from above, and the right-hand side of (30) will grow without bound. This means that there exists an equilibrium for large values of λ_E/λ_F . The right-hand side of (30) converges to zero as ρ grows without bound. Thus, an equilibrium exists for all large enough ρ .

B. The Large λ_E/λ_F Asymptote

The right-hand side of the zero-profit condition can only grow without bound if $r - \kappa$ approaches zero or ζ approaches 1. If $\eta > 0$, then $r - \kappa$ must be positive and bounded away from zero since $r > \kappa + \eta$ in any equilibrium. If $\eta = 0$, then $\zeta \geq 1$ implies that μ is negative and bounded away from zero. In that case, $\xi/(r - \kappa)$ converges to $1/|\mu| < \infty$ if $r - \kappa$ approaches zero from above. This means that the right-hand side of (30) can only grow without bound as ζ approaches 1 from above.

APPENDIX II

Proof of Proposition 3. The survivor function of a cohort of firms entering at the same time is the average of the conditional survivor function $\Lambda(a|x)$ based on the size distribution of successful entrants. The density of this distribution is proportional to $f(x + \delta)$ and can be written as a weighted average of the exponential density $f_e(x) = \zeta e^{-\zeta(x-b)}$ and the gamma density $f_g(x) = \zeta^2(x - b)e^{-\zeta(x-b)}$. Calculating the appropriate weights gives

$$\Lambda(a) = \left(\frac{\delta\zeta}{1 + \delta\zeta} \right) \Lambda_e(a) + \left(\frac{1}{1 + \delta\zeta} \right) \Lambda_g(a),$$

where $\Lambda_e(a)$ is the survivor function based on initial conditions drawn from $f_e(x)$, and $\Lambda_g(a)$ is the survivor function based on initial conditions drawn from $f_g(x)$. The resulting hazard rate $h(a) = -D\Lambda(a)/\Lambda(a)$ is a weighted average of the hazard rates $h_e(a) = -D\Lambda_e(a)/\Lambda_e(a)$ and $h_g(a) = -D\Lambda_g(a)/\Lambda_g(a)$.

Define $\Psi(x) = x\Phi(-x)/\phi(x)$ and write $u = (-\mu/\sigma)\sqrt{a}$ and $v = ([\mu + \zeta\sigma^2]/\sigma)\sqrt{a}$. The survivor functions for exponential and gamma initial conditions are

$$\Lambda_e(a) = \frac{\mu}{\mu + (1/2)\sigma^2\zeta} [\Psi(u) - \Psi(v)] \frac{1}{u} \phi(u),$$

and

$$\Lambda_g(a) = \frac{\mu(\mu + \zeta\sigma^2)}{(\mu + (1/2)\zeta\sigma^2)^2} \left[\Psi(u) - \Psi(v) - \frac{u^2 - v^2}{2v^2} (1 - (1 + v^2)[1 - \Psi(v)]) \right] \frac{\phi(u)}{u},$$

respectively. The resulting hazard rates can be written as

$$h_e(a) = \frac{u^2 - v^2}{2a} \frac{1 - \Psi(v)}{\Psi(u) - \Psi(v)},$$

and

$$h_g(a) = \frac{u^2 - v^2}{2a} \left[1 + \frac{\Psi(v) - \Psi(u)}{(1 - (1 + v^2)[1 - \Psi(v)])(u^2 - v^2)/(2v^2)} \right]^{-1}.$$

If $\delta > 0$, then both $h_e(a)$ and $h_g(a)$ are decreasing, and this implies that $h(a)$ is decreasing. To prove that $h_e(a)$ and $h_g(a)$ are decreasing, one can use continued-fraction upper and lower bounds for the Mill's ratio $\Phi(-x)/\phi(x)$ reported in Lee [1992]. These bounds can also be used to establish the asymptote reported in Proposition 3. A lengthy proof is available at www.luttmer.org. At $\delta = 0$, the hazard rate is constant because (23) implies that the stationary density is an eigenfunction of the operator $-\mu Df(x) + \sigma^2 D^2f(x)/2$.

UNIVERSITY OF MINNESOTA
FEDERAL RESERVE BANK OF MINNEAPOLIS

REFERENCES

Aghion, Philippe, and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, LX (1992), 323-351.

- Aghion, Philippe, Christopher Harris, Peter Howitt, and John Vickers, "Competition, Imitation and Growth with Step-by-Step Innovation," *Review of Economic Studies*, LXVIII (2001), 467–492.
- Arrow, Kenneth J., "The Economic Implications of Learning-by-Doing," *Review of Economic Studies*, XXIX (1962), 155–173.
- Atkeson, Andrew, and Patrick J. Kehoe, "Modeling and Measuring Organization Capital," *Journal of Political Economy*, CXIII (2005), 1026–1053.
- Audretsch, David B., "New-Firm Survival and the Technological Regime," *The Review of Economics and Statistics*, LXXIII (1991), 441–450.
- Axtell, Robert L., "Zipf Distribution of U. S. Firm Sizes," *Science*, CCXCIII (2001), 1818–1820.
- Barro, Robert J., and Xavier Sala-i-Martin, *Economic Growth*, 2nd ed. (Cambridge, MA: MIT Press, 2004).
- Boldrin, Michele, and David K. Levine, "Perfectly Competitive Innovation," University of Minnesota Working Paper, 2000.
- Boldrin, Michele, and Jose A. Scheinkman, "Learning-by-Doing, International Trade and Growth: A Note," in *The Economy as an Evolving Complex System*, Philip W. Anderson, Kenneth J. Arrow, and David Pines, eds. (Reading, MA: Addison-Wesley, 1988).
- Brüderl, Josef, Peter Preisdörfer, and Rolf Ziegler, "Survival Chances of Newly Founded Business Organizations," *American Sociological Review*, LVII (1992), 227–242.
- Cabral, Luis M. B., and José Mata, "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, XCIII (2003), 1075–1090.
- Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu, "Have Industrial Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, LVI (2001), 1–43.
- Caves, Richard E., "Industrial Organization and New Findings on the Turnover and Mobility of Firms," *Journal of Economic Literature*, XXXVI (1998), 1947–1982.
- Dixit, Avinash K., and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, LXVII (1977), 297–308.
- Dixit, Avinash K., and Robert S. Pindyck, *Investment under Uncertainty* (Princeton, NJ: Princeton University Press, 1994).
- Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, "Patterns of Firm Entry and Exit in U. S. Manufacturing Industries," *RAND Journal of Economics*, IX (1988), 495–515.
- , "The Growth and Failure of U. S. Manufacturing Plants," *Quarterly Journal of Economics*, CIV (1989), 671–698.
- Eaton, Jonathan, and Zvi Eckstein, "Cities and Growth: Theory and Evidence from France and Japan," *Regional Science and Urban Economics*, XXVII (1997), 443–474.
- Eeckhout, Jan, "Gibrat's Law for (All) Cities," *American Economic Review*, XCIV (2004), 1429–1451.
- Eeckhout, Jan, and Boyan Jovanovic, "Knowledge Spillovers and Inequality," *American Economic Review*, XCII (2002), 1290–1307.
- Feller, William, *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed. (New York, NY: Wiley, 1971).
- Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, CXIV (1999), 739–767.
- Gabaix, Xavier, and Yannis M. Ioannides, "The Evolution of City Size Distributions," in *Handbook of Regional and Urban Economics*, vol IV: *Cities and Geography*, J. V. Henderson, and Jean-Francois Thisse, eds. (Amsterdam, The Netherlands: Elsevier, 2003).
- Gibrat, Robert, *Les Inégalités Économiques* (Paris, France: Librairie du Recueil Sirey, 1931).
- Grossman, Gene M., and Elhanan Helpman, *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press, 1991).
- Harrison, J. Michael, *Brownian Motion and Stochastic Flow Systems* (New York, NY: Wiley, 1985).
- Headd, Brian, "Redefining Business Success: Distinguishing between Closure and Failure," *Small Business Economics*, XXI (2003), 51–61.

- Hellwig, Martin F., and Andreas Irmen, "Endogenous Technical Change in a Competitive Economy," *Journal of Economic Theory*, CI (2001), 101, 1–39.
- Hopenhayn, Hugo, "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, LX (1992), 1127–1150.
- Ijiri, Yuji, and Herbert A. Simon, "Business Firm Growth and Size," *American Economic Review*, LIV (1964), 77–89.
- Jones, Larry E., and Rodolfo E. Manuelli, "A Convex Model of Equilibrium Growth: Theory and Policy Implications," *Journal of Political Economy*, XCVIII (1990), 1008–1038.
- Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, C (1982), 649–670.
- , and Glenn M. MacDonald, "Competitive Diffusion," *Journal of Political Economy*, CII (1994), 24–52.
- Karlin, Samuel, and Howard M. Taylor, *A Second Course in Stochastic Processes* (San Diego, CA: Academic Press, 1981).
- Klette, Tor J., and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, CXII (2004), 986–1018.
- Krugman, Paul, "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, IX (1979), 469–480.
- Lee, Chu-In C., "On Laplace Continued Fraction for the Normal Integral," *Annals of the Institute of Statistical Mathematics*, XCIV (1992), 107–120.
- Lucas, Robert E. Jr., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, IX (1978), 508–523.
- , "On the Mechanics of Economic Development," *Journal of Monetary Economics*, XXII (1988), 3–42.
- Luttmer, Erzo G. J., "The Size Distribution of Firms in an Economy with Fixed and Entry Costs," Federal Reserve Bank of Minneapolis Working Paper No. 633, 2004.
- Mata, José, and Pedro Portugal, "Life Duration of New Firms," *The Journal of Industrial Organization*, XLII (1994), 227–245.
- Melitz, Marc, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, CXXI (2003), 1695–1725.
- Miao, Jianjun, "Optimal Capital Structure and Industry Dynamics," *Journal of Finance*, LX (2005), 2621–2659.
- Nelson, Richard R., and Sidney G. Winter, *An Evolutionary Theory of Economic Change* (Cambridge, MA: Belknap Press, 1982).
- Nucci, Alfred R., "The Demography of Business Closings," *Small Business Economics*, XII (1998), 25–39.
- Pakes, Ariel, and Richard Ericson, "Empirical Implications of Alternative Models of Firm Dynamics," *Journal of Economic Theory*, LXXIX (1998), 1–45.
- Parente, Stephen L., and Edward C. Prescott, "Monopoly Rights: A Barrier to Riches," *American Economic Review*, LXXXIX (1999), 1216–1233.
- Romer, Paul, "Endogenous Technological Change," *Journal of Political Economy*, XCVIII (1990), 71–102.
- Rossi-Hansberg, Esteban, and Mark L. J. Wright, "Firm Size Dynamics in the Aggregate Economy," Stanford University, Palo Alto, CA, 2004.
- Segerstrom, Paul S., "Innovation, Imitation, and Economic Growth," *Journal of Political Economy*, XCIX (1991), 807–827.
- Simon, Herbert A., and Charles P. Bonini, "The Size Distribution of Business Firms," *American Economic Review*, XCVIII (1958), 607–617.
- Steindl, Josef, *Random Processes and the Growth of Firms; A Study of the Pareto Law* (New York, NY: Hafner Publishing Company, 1965).
- Stokey, Nancy L., "Learning-by-Doing and the Introduction of New Goods," *Journal of Political Economy*, XCVI (1988), 701–717.
- Sutton, John, "Gibrat's Legacy," *Journal of Economic Literature*, XXXV (1997), 40–59.
- , "The Variance of Firm Growth Rates: The 'Scaling' Puzzle," *Physica A*, CCCXII (2002), 577–590.
- Young, Alwyn, "Learning by Doing and the Dynamic Effects of International Trade," *Quarterly Journal of Economics*, CVI (1991), 369–405.