Federal Reserve Bank of Minneapolis
Research Department

Slow Convergence in Economies
with Firm Heterogeneity[*]

Erzo G.J. Luttmer

Working Paper 696

Revised March 2012

ABSTRACT _____

This paper presents a simple formula that relates the tail index of the firm size distribution to
the aggregate speed with which an economy converges to its balanced growth path. The fact that
there are so many firms in the right tail implies that aggregate shocks that permanently destroy
employment among incumbent firms, rather than cause these firms to scale back temporarily, are
followed by slow recoveries. This is true despite the presence of many rapidly growing firms. Aggre-
gate convergence rates are non-linear: they can be very high for economies far below the balanced
growth path and very low for advanced economies.

JEL classification: E1, L1
Keywords: Firm growth; Firm size distribution; Recessions; Recoveries

1. Introduction

What happens when an economy is hit by an aggregate shock that destroys a significant amount of employment? Reinhart and Rogoff [2008] have argued that financial crises lead to recessions followed by slow recoveries. Both the initial downturns and the subsequent slow recoveries are routinely interpreted using models that rely on credit market imperfections. A credit crunch can take out firms, and if new and surviving firms have to finance growth internally, then slow growth of internally generated funds can lead to slow growth of firms and aggregate output.

The competing hypothesis is that there are other shocks that cause recessions, and that financial crises are symptomatic of large shocks. Any attempt to make this case requires an alternative large aggregate shock, an account of the phenomena that seem to be typical of financial crises, and an alternative explanation for the slow subsequent recoveries. This paper addresses this last piece of the puzzle by showing that slow recoveries are a built-in feature of some prototypical models of firm heterogeneity that do not involve credit market imperfections.

One striking aspect of firm heterogeneity is simply the great variation in firm size. The distribution of employment across firms is approximately Pareto, and the tail of this Pareto distribution is so heavy that the distribution almost does not have a finite mean. The right tail behaves like $x^{-\zeta}$ with $\zeta \approx 1.05$, and the mean of a Pareto distribution is no longer finite at $\zeta = 1$. The rate at which aggregate output converges to its balanced growth path is examined here for two different economies that can account for this type of firm heterogeneity.

In the first economy, firms grow because of post-entry productivity growth, at a constant rate that is taken to be exogenous. In this economy, the speed of convergence of aggregate output is approximately

$$(\eta + \delta)(1 - 1/\zeta)(1 + \mathcal{E}). \tag{1}$$

In this formula, η is the population growth rate, δ is the rate at which firms exit randomly, and $\mathcal{E}$ is the elasticity of firm entry with respect to a q-like variable—specifically, the value of a new firm scaled by per-capita consumption. For example, if $\eta = 0.01$, $\delta = 0.10$ and $\mathcal{E} = 2$, then the above estimate for ζ gives a speed of convergence equal to 0.016, implying a half-life of about 44 years. Slow growth of incumbent firms is not the reason for this slow convergence: in the economy that gives rise to this calculation, surviving firms grow at a rate $(\eta + \delta)/\zeta$, or about 10.5% per annum.

There are four ingredients that give rise to this result. First, it takes a long time to

1

build up the large firms that account for so much of aggregate employment. As reported in Luttmer [2011a], the median firm with 10,000 or more employees is about 70 years old. Second, firms become large by growing faster than average for some time. In the above calculation, surviving firms grow at a 10.5% rate while aggregate employment grows at 1% per annum. Third, while the evidence is still scant, it appears to be the case that firms on a path of rapid growth do not significantly adjust their rates of expansion across stages of the business cycle. This is illustrated in Figure 1, which shows the employment histories of some rapidly growing US firms, together with US unemployment as a business cycle indicator.[1] Fourth, there is only so much new entrants can contribute to output and employment. Entrants are small, and it would take very high entry rates to compensate for a large negative shock.
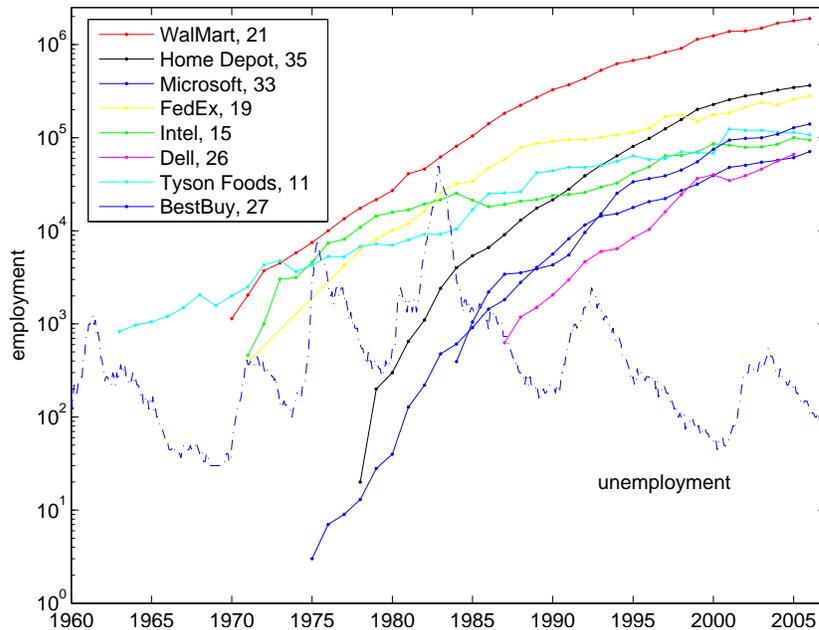

FIGURE 1 Rapid Firm Growth and the Business Cycle
(annual firm growth rates in %)

The model for this first economy is a drastically stripped-down version of Luttmer [2007] in which incumbent firms grow at a constant rate conditional on survival, as a result of exogenous productivity growth. As in Luttmer [2011a], agents differ in their abilities to

---

[1]See Moscarini and Postel-Vinay [2011] for evidence on the contributions of large and small employers to employment growth at different stages of the business cycle. Haltiwanger, Jarmin, and Miranda [2010] emphasize the importance of firm age. In the economy of Luttmer [2011a], age will proxy for an underlying state variable that characterizes rapidly growing firms. The hypothesis here is that this state variable is important for predicting how firms will behave over the business cycle.

supply labor and set up new firms, and new firms are small. Comparative advantage and the state of the economy determine labor supply and the rate of entry, giving rise to the elasticity $\mathcal{E}$ used in the above calculation.

In the second economy, firm growth is not driven by technical change. Rather, firms can use labor inputs not only to produce consumption, but also to replicate the projects that allow them to produce particular differentiated goods. New projects are created not only from scratch when entrepreneurs set up new firms, but also using a technology with existing projects and labor as inputs. Suppose this technology exhibits constant returns to scale. Let $\mathcal{S}$ be the factor share of labor in this replication process and write $\mathcal{F}$ for the curvature of the production function for new projects, holding fixed the input of existing projects and varying the input of labor. If existing projects fail at the rate λ, consumers discount flow utilities at a positive ρ, and the Frisch elasticity of labor supply is infinite, as in Hansen [1985] and Rogerson [1988], then the speed of convergence in this economy is

$$\frac{\rho}{2}\left(-1 + \sqrt{1 + 4\left(1 + \frac{(\eta+\lambda)\mathcal{S}}{\rho}\right)\frac{(\eta+\lambda)\mathcal{S}}{\rho\mathcal{F}}}\right), \qquad (2)$$

when the size distribution of firms has a tail index $\zeta \downarrow 1$. In (1), setting $\zeta = 1$ would imply no convergence at all. But here the incumbent replication process can respond to the state of the economy. Because of this, mean reversion rates in the aggregate economy remain bounded away from zero even as $\zeta \downarrow 1$.[2] For example, suppose the subjective discount rate is $\rho = 0.02$ and projects fail at the rate $\lambda = 0.24$, a rate that is consistent with the high observed exit rates of small firms. Then a population growth rate of $\eta = 0.01$ gives $(\eta+\lambda)/\rho = 12.5$. If the technology for replicating projects is Cobb-Douglas, then $\mathcal{F} = 1 - \mathcal{S}$ and a factor share of $1/2$ implies a convergence rate of about 0.18, corresponding to a half-life of a bit more than 3.8 years. This is an order of magnitude faster than the half-lives implied by (1). The Cobb-Douglas assumption is critical for this relatively rapid rate of convergence. If the elasticity of substitution between existing projects and labor in producing new projects is significantly below 1, then $\mathcal{F}$ can be well above the factor share of existing projects. For example, a constant elasticity of $1/3$ implies $\mathcal{F} = 3(1 - \mathcal{S})$, and at a factor share $\mathcal{S} = 1/2$ this reduces (2) to about 0.10, resulting in a half life of almost 7 years. This speed of convergence is less extreme than what is implied by (1), but slow enough to make it easy to account for the

---

[2] Atkeson and Burstein [2010] present an example of slow convergence in an economy in which firms can improve their own productivity. This should also keep convergence rates bounded away from zero as the tail index of the firm size distribution converges to 1.

slow recoveries seen after recent US recessions.

The rest of this paper proceeds as follows. Section 2 sets up the entirely familiar structure of product markets with monopolistically competitive producers. Section 3 describes entry and incumbent productivity growth and shows the two types of balanced growth paths that can arise: one driven by incumbent productivity growth in which the firm size distribution is determined by initial conditions, and another in which entry productivity drives aggregate growth. The latter is the economy that gives rise to a stationary firm size distribution that is Pareto, and to the rate of convergence calculation reported above. Although there are two physical state variables, the economy has a recursive structure that allows one to use a simply phase diagram to describe the equilibrium in this economy. Section 4 constructs this phase diagram, determines the speed of convergence shown in (1), and shows how slow convergence near the balanced growth path is consistent with rapid growth in economies that are well below the balanced growth path. Section 5 describes the alternative economy that gives rise to (2). Section 6 concludes and describes an important extension.

2. CONSUMERS AND PRODUCT MARKETS

The population is $H_t = He^{\eta t}$. The per-capita endowment of labor is 1, but some agents may choose to become entrepreneurs rather than supply labor as workers. Preferences over composite consumption are logarithmic, and the subjective discount rate is $\rho > 0$. Markets are complete and so aggregate consumption flows can be used to calculate present values. Composite consumption flows C_t imply discount factors $e^{-\rho t}/C_t$. With these discount factors, calculating present values boils down scaling payoff by aggregate consumption and discounting the resulting scaled payoffs at the rate ρ. The implied interest rates are $r_t = \rho + \mathrm{D}C_t/C_t$. Note well that C_t is aggregate consumption, and ρ should be interpreted as the effective subjective discount rate of a growing dynasty of consumers.

2.1 Differentiated Commodities

The composite good at time t is

$$C_t = \left(\int (C_t[z])^{1-1/\varepsilon} \, \mathrm{d}M_t[z] \right)^{1/(1-1/\varepsilon)},$$

where $\varepsilon > 1$ and $M_t[z]$ represents the measure of type-z goods. For simplicity, all goods are equally weighted. Since $\varepsilon > 1$, unavailable goods do not affect the composite good.

The type of a good will be defined by the technology for producing it. Assume $\varepsilon > 1$ so that zero consumption of a particular good is consistent with positive composite consumption. The price of type-z goods is $p_t[z]$, in some arbitrary unit of account. The demand curves are

$$\frac{C_t[z]}{C_t} = \left(\frac{p_t[z]}{P_t}\right)^{-\varepsilon} \tag{3}$$

where

$$P_t = \left(\int (p_t[z])^{1-\varepsilon}\,\mathrm{d}M_t[z]\right)^{1/(1-\varepsilon)}$$

is the price index for the composite good.

2.2 Type-z Producers

There is only one producer for every good. The producer of a type-z good has access to a technology that is linear in labor, and z represents labor productivity. Let w_t be the wage in units of the composite good. The marginal cost of producing good z is therefore w_t/z in units of the composite good. Producers set prices and produce whatever is demanded. The familiar markup formula gives

$$\frac{p_t[z]}{P_t} = \frac{1}{1 - 1/\varepsilon}\frac{w_t}{z}. \tag{4}$$

Using this to eliminate the $p_t[z]/P_t$ from the definition of the price index gives

$$w_t = (1 - 1/\varepsilon)Z_t N_t^{1/(\varepsilon-1)}, \tag{5}$$

where

$$N_t = \int \mathrm{d}M_t[z], \quad Z_t = \left(\frac{1}{N_t}\int z^{\varepsilon-1}\mathrm{d}M_t[z]\right)^{1/(\varepsilon-1)}, \tag{6}$$

provided that the measure $M_t[\cdot]$ is such that both N_t and Z_t are well defined. The fact that prices behave like $p_t[z] \propto 1/z$ implies that the price index P_t is scales with an average of $z^{\varepsilon-1}$, and so none of this is well defined if the mean of $z^{\varepsilon-1}$ is not finite.

The real wage is determined the measure of goods N_t and the productivity index Z_t, just as in a competitive economy. The real wage is high when the aggregate productivity index Z_t is high, and when there are many goods. It is lower than it would be in a competitive economy by a factor $1 - 1/\varepsilon$. The relative prices $p_t[z]/P_t$ are exactly the same as in a competitive economy.

2.3 The Demand for Labor

Let $L_t \in [0, H_t]$ be the amount of labor supplied and write $L_t[z]$ for the amount of labor used to produce a specific type-z good (note that all the type-z producers do the same thing.) The labor market clearing condition is

$$\int L_t[z] \mathrm{d} M_t[z] = L_t.$$

Using $L_t[z] = C_t[z]/z$, the demand curves for $C_t[z]/C_t$, the equilibrium relative prices $p_t[z]/P_t$ given in (4), and the wages w_t given in (5), one can use this labor market clearing condition to solve for C_t. The result is

$$C_t = Z_t N_t^{1/(\varepsilon-1)} L_t. \tag{7}$$

This implies $C_t = w_t L_t/(1 - 1/\varepsilon)$, reflecting the fact that all producers use the same markup over marginal cost. Note that $p_t[z] \propto 1/z$ and so $C_t[z] = L_t[z]/z$. Expenditure shares and employment shares must be the same. They are given by

$$\frac{p_t[z] C_t[z]}{P_t C_t} = \frac{L_t[z]}{L_t} = \frac{1}{N_t} \left(\frac{z}{Z_t} \right)^{\varepsilon-1}. \tag{8}$$

The definition of Z_t shows that these shares do indeed aggregate to 1. If all producers have the same labor productivity, then these shares are simply $1/N_t$. If labor productivity is heterogeneous, then more productive producers will use more labor and have higher sales.

2.4 Profits

Write $v_t[z]$ for the profits earned by the producer of a type-z good, measured in units of the composite good. These profits are $v_t[z] = (p_t[z]/P_t - w_t/z) C_t[z]$. The markup formula gives $v_t[z] = (1/\varepsilon) p_t[z] C_t[z]/P_t$, and then the expenditure shares (8) imply

$$\frac{v_t[z]}{C_t} = \frac{1}{\varepsilon} \frac{1}{N_t} \left(\frac{z}{Z_t} \right)^{\varepsilon-1}.$$

Thus profit shares are the same as expenditure shares and employment shares. Aggregate profits are obtained by integrating against $M_t[z]$, and this confirms that aggregate profits are a fraction $1/\varepsilon$ of aggregate consumption. Since $w_t L_t = (1 - 1/\varepsilon) C_t$, one can write

$$\frac{v_t[z]}{w_t} = \frac{1}{\varepsilon - 1} \left(\frac{z}{Z_t} \right)^{\varepsilon-1} \frac{L_t}{N_t}. \tag{9}$$

Thus profits scale with labor supply per firm, and their distribution across producers follows from the distribution of labor productivities.

3. Technology Adoption

Entrants can adopt technologies with labor productivities that are exogenously improving over time. Entry productivity is given by

$$X_t = X e^{\theta_E t},$$

where θ_E is positive. Following entry, incumbent productivity grows at the constant rate θ_I until a shock, arriving randomly at the rate $\delta > 0$, hits. This shock causes productivity to drop to zero, permanently. Write $z_{t,a}$ for the time-$t + a$ productivity of a surviving firm that entered at time t. Then $z_{t,a} = X_t e^{\theta_I a}$, and hence

$$\frac{z_{t-a,a}}{X_t} = e^{(\theta_I - \theta_E)a}. \tag{10}$$

That is, surviving incumbent productivity relative to current entry productivity is a function only of incumbent age. It is easy to modify this very simple process of incumbent productivity in a way that gives rise to rapidly growing young firms that transition with some probability to more slowly growing mature firms, as in Luttmer [2011a]. Most of the analytical tractability of the current setup is maintained.

There are now two sources of technological progress. Depending on parameters, one or the other will determine the aggregate growth rate of the economy. Which one it is has implications for how the distribution of employment across firms evolves over time.

3.1 No Entry

Suppose initially that no entry is possible. Nobody becomes and entrepreneur and hence the per-capita labor supply is 1. The aggregate labor supply is thus $L_t = H_t$. It follows from (7) that aggregate consumption is given by $C_t = Z_t N_t^{1/(\varepsilon-1)} H_t$. The measure of firms shrinks over time according to $DN_t = -\delta N_t$ and the productivity index follows

$$D\left[N_t Z_t^{\varepsilon-1}\right] = [(\varepsilon - 1)\theta_I - \delta]N_t Z_t^{\varepsilon-1}.$$

The first term comes from the incumbent productivities that grow at the rate θ_I and the second from the fact that firms disappear at the rate δ. Since $C_t/H_t = Z_t N_t^{1/(\varepsilon-1)}$, this implies

$$\frac{D[C_t/H_t]}{C_t/H_t} = \theta_I - \frac{\delta}{\varepsilon - 1}. \tag{11}$$

Incumbent productivity improvements add to per-capita consumption growth. But the loss of producers reduces variety, and this reduces per-capita consumption growth.

Any initial heterogeneity in productivity in this economy persists over time. Everyone trends up at the same rate θ_I. Incumbents exit randomly, and not because they have low productivity. The distribution of productivities scaled by $e^{-\theta_I t}$ remains what it was at the initial date.

3.2 Costly Entry

Suppose the value of a new firm is q_t, measured in units of the composite good. The creation of new firms requires entrepreneurial effort. Suppose that agents are heterogeneous in terms of their ability to be entrepreneurs. Let y be the average flow of entry opportunities an agent can create per unit of time. Let x be the flow of labor services the agent can supply. Then this agent will choose to be an entrepreneur if $q_t y > w_t x$, and a worker if the reverse inequality holds. Thus individuals make their labor supply decisions based on q_t/w_t. Given a talent distribution T over (x, y), this yields per-capita supplies of entrepreneurial effort and labor equal to

$$\begin{bmatrix} L(q_t/w_t) \\ E(q_t/w_t) \end{bmatrix} = \int \begin{bmatrix} x \iota\{q_t y < w_t x\} \\ y \iota\{q_t y > w_t x\} \end{bmatrix} \mathrm{d}T(x, y)$$

and hence

$$E_t = E(q_t/w_t)H_t, \quad L_t = L(q_t/w_t)H_t. \tag{12}$$

These factor supplies are well defined if the talent distribution has a finite mean, and one can normalize the units in which labor is measured to ensure that the maximum per-capita labor supply is $L(0) = 1$. These factor supplies are smooth if the talent distribution has a density, and positive at all positive q_t/w_t if the talent distribution has full support in $(0, \infty) \times (0, \infty)$. Clearly, the function $E(\cdot)$ will be increasing and $L(\cdot)$ decreasing. The higher the value of a new firm relative to wages, the more agents will choose to be entrepreneurs, and the fewer to be workers. As noted in Luttmer [2011a], if the talent distribution is independent Fréchet, then the elasticity of $E(q/w)/L(q/w)$ will be constant.

3.2.1 The Value of a New Firm

It will be convenient to relate the value of a new firm to the aggregate value of all firms. Let $q_t[z]$ be the value of a firm with current productivity z. Then the aggregate value of all firms is given by

$$V_t = \int q_t[z]\mathrm{d}M_t[z].$$

By (9), the profits of a type-z firm scale with $z^{\varepsilon-1}$. Since new firms enter with productivity X_t, it follows that $v_t[z] = (z/X_t)^{\varepsilon-1}v_t[X_t]$. The labor productivities for all surviving firms grow at the rate θ_{I}. This then implies that the value of a firm with productivity z at time t is given by

$$q_t[z] = \left(\frac{z}{X_t}\right)^{\varepsilon-1} q_t. \tag{13}$$

It follows that

$$V_t = q_t S_t$$

where

$$S_t = \int \left(\frac{z}{X_t}\right)^{\varepsilon-1} \mathrm{d}M_t[z]. \tag{14}$$

Note that $S_t = N_t(Z_t/X_t)^{\varepsilon-1}$, and thus the productivity index Z_t can be inferred from S_t, X_t and N_t. The variable S_t will be a more convenient state variable to use than Z_t. Note that S_t would equal N_t if there was no heterogeneity in productivities.

Recall that $w_t L_t = (1 - 1/\varepsilon)C_t$, and thus

$$\frac{q_t}{w_t} = \frac{1}{1 - 1/\varepsilon}\frac{L_t}{S_t}\frac{V_t}{C_t} \tag{15}$$

The variable S_t is a state variable, and $L_t = L(q_t/w_t)H_t$ is itself be a function of q_t/w_t because of the occupational choices made by agents in this economy. The aggregate value of all firms depends on beliefs about future profits. The same must then be true for V_t/C_t, and the logarithmic utility assumption makes this a convenient forward-looking state variable to use.

3.3 The Dynamics of N_t and S_t

The number of producers evolves according to

$$\mathrm{D}N_t = -\delta N_t + E_t \tag{16}$$

where E_t is given as a function of H_t and q_t/w_t in (12). The state variable is an integral of $(z/X_t)^{\varepsilon-1}$ against the measure $M_t[z]$. For new entrants, this integrand equals 1. Incumbents exit randomly at the rate δ. Together with the fact that X_t grows at the rate θ_{E} and the productivity of surviving incumbent producers at the rate θ_{I}, this implies

$$\mathrm{D}S_t = [(\varepsilon - 1)(\theta_{\mathrm{I}} - \theta_{\mathrm{E}}) - \delta]\,S_t + E_t. \tag{17}$$

The key observation here is that we can trace the evolution of the state variable S_t without keeping track of the entire cross-sectional distribution of productivity.

9

3.4 The Dynamics of V_t

It remains to construct a differential equation for V_t/C_t. Recall $z_{t,a}$ is the productivity of a time-t entrant at time $t+a$. Thus $z_{s,t-s}$ is the productivity at time t of a time-$(t-s)$ entrant. The value of such a firm is $q_t[z_{s,t-s}]$, and this value must satisfy the usual asset pricing equation,

$$(r_t + \delta)q_t[z_{s,t-s}] = v_t[z_{s,t-s}] + \frac{\mathrm{d}}{\mathrm{d}t}q_t\,[z_{s,t-s}]\,.$$

Now note from (9), (10) and (13) that

$$\begin{bmatrix} v_t[z_{s,t-s}] \\ q_t[z_{s,t-s}] \end{bmatrix} = e^{(\varepsilon-1)(\theta_\mathrm{I}-\theta_\mathrm{E})(t-s)} \begin{bmatrix} v_t[X_t] \\ q_t \end{bmatrix}.$$

Combining this with the above asset pricing equation gives

$$(r_t + \delta)q_t = v_t[X_t] + (\varepsilon - 1)(\theta_\mathrm{I} - \theta_\mathrm{E})q_t + \mathrm{D}q_t.$$

Multiplying by S_t and using $V_t = q_t S_t$ and (17) shows that

$$\mathrm{D}V_t = r_t V_t + q_t E_t - v_t[X_t]S_t. \tag{18}$$

The aggregate value of all firms grows as the value of existing firms grows at the rate of interest, by more to an extent determined by the investment that occurs at entry, and by less to an extent determined by the distribution of profits.

The logarithmic utility assumption implies $r_t = \rho + \mathrm{D}C_t/C_t$. The value of entering firms is $q_t = V_t/S_t$, and the fact that monopolist producers earn a fraction $1/\varepsilon$ of their revenues implies that aggregate profits are $v_t[X_t]S_t = C_t/\varepsilon$. With this, (18) can be written as

$$\mathrm{D}\left[\frac{V_t}{C_t}\right] = \left(\rho + \frac{E_t}{S_t}\right)\frac{V_t}{C_t} - \frac{1}{\varepsilon}. \tag{19}$$

Consumer wealth consists of labor income and the aggregate value of firms. Both are non-negative, and thus the transversality condition for the representative dynastic consumer implies

$$\lim_{t\to\infty} e^{-\rho t}\left[\frac{V_t}{C_t}\right] = 0. \tag{20}$$

Solving (19) forward and using (20) shows that V_t/C_t is the present value of $1/\varepsilon$ calculated using the discount rates $\rho + E_t/S_t \geq \rho > 0$. As required, this will be well defined and positive.

3.5 The Equilibrium

We now have differential equations (16), (17) and (19) for N_t, S_t, and V_t/C_t, together with initial conditions for N_0 and S_0, and the transversality condition (20). These differential equations depend on E_t and q_t/w_t, and these variables must satisfy (12) and (15). The labor-supply function (12) and the expression (15) for q_t/w_t imply

$$\frac{q_t}{w_t} = \frac{L(q_t/w_t)}{1 - 1/\varepsilon} \frac{V_t/C_t}{S_t/H_t}. \tag{21}$$

The differential equation (17) for S_t and the entrepreneurial effort supply (12) imply

$$D\left[\frac{S_t}{H_t}\right] = -\left[\eta + \delta - (\varepsilon - 1)(\theta_{\mathrm{I}} - \theta_{\mathrm{E}})\right]\left[\frac{S_t}{H_t}\right] + E(q_t/w_t) \tag{22}$$

The differential equation (18) together with (12) becomes

$$D\left[\frac{V_t}{C_t}\right] = \left(\rho + \frac{E(q_t/w_t)}{S_t/H_t}\right)\left[\frac{V_t}{C_t}\right] - \frac{1}{\varepsilon} \tag{23}$$

and we have the transversality condition

$$\lim_{t \to \infty} e^{-\rho t}\left[\frac{V_t}{C_t}\right] = 0. \tag{24}$$

Note that (21) pins down q_t/w_t in terms of S_t/H_t and V_t/C_t. Thus the equations (21)-(23) define a differential equation for S_t/H_t and V_t/C_t. The value of S_0/H_0 is an initial condition determined by how many firms with various productivity levels there are at the initial date, and (24) is a boundary condition at infinity.

The equilibrium conditions (21)-(24) determine the ratios S_t/H_t, V_t/C_t and q_t/w_t. From (7) and the definition (14) of S_t, per-capita consumption can be written as

$$\frac{C_t}{H_t} = X_t H_t^{1/(\varepsilon-1)} \left(\frac{S_t}{H_t}\right)^{1/(\varepsilon-1)} L(q_t/w_t). \tag{25}$$

This determines the equilibrium path of consumption, and wages then follow from $(1 - 1/\varepsilon)C_t = w_t H_t L(q_t/w_t)$. The equilibrium path for q_t/w_t together with (16) can be used to compute the path for N_t.

3.6 Long-Run Growth Rates

The factor $X_t H_t^{1/(\varepsilon-1)}$ in (25) grows at the rate $\theta_{\mathrm{E}} + \eta/(\varepsilon - 1)$ and the per-capita supply of labor is bounded. So the long-run growth rate of this economy depends critically on

11

what happens to S_t/H_t. In turn, this depends on whether or not the parameters happen to satisfy

$$\eta + \delta > (\varepsilon - 1)(\theta_{\mathrm{I}} - \theta_{\mathrm{E}}). \tag{26}$$

If this inequality holds, then the sign of the coefficient on S_t/H_t in (22) is negative, and otherwise it is positive.

3.6.1 Vanishing Entry

Suppose first that the inequality (26) is reversed. The supply of entrepreneurial effort $E(u_t)$ is bounded above by the entrepreneurial effort that would result if everyone became an entrepreneur. The differential equation (22) then implies that S_t/H_t will asymptotically grow at the rate $-[(\eta + \delta) - (\varepsilon - 1)(\theta_{\mathrm{I}} - \theta_{\mathrm{E}})] > 0$. Eventually, $E(u_t)/(S_t/H_t)$ will be negligible, and so (23)-(24) implies that V_t/C_t converges to $1/(\rho\varepsilon)$. The asymptotic growth rate of S_t/H_t together with (25) implies that per-capita consumption grows at the rate

$$\frac{\mathrm{D}[C_t/H_t]}{C_t/H_t} = \theta_{\mathrm{I}} - \frac{\delta}{\varepsilon - 1} \tag{27}$$

in the long run. The economy is essentially the same as the one without entry discussed earlier. Its growth rate is entirely determined by what happens to the productivity and number of incumbent producers.

3.6.2 Persistent Entry

Suppose now that the parameters do happen to satisfy (26) and conjecture that $E(q_t/w_t)$ converges to a positive constant. Then (17) implies that S_t/H_t converges to a positive constant, at the positive rate $\eta + \delta - (\varepsilon - 1)(\theta_{\mathrm{I}} - \theta_{\mathrm{E}})$. The conjectured convergence of $E(q_t/w_t)$ also means that $L(q_t/w_t)$ converges, and so per-capita consumption will grow at the rate

$$\frac{\mathrm{D}[C_t/L_t]}{C_t/L_t} = \theta_{\mathrm{E}} + \frac{\eta}{\varepsilon - 1}. \tag{28}$$

in the long run. That is, the growth rate of per-capita consumption is determined by entry productivity growth and gains from variety that arise because the number of producers grows with the population. Note that (26) can be written as

$$\theta_{\mathrm{E}} + \frac{\eta}{\varepsilon - 1} > \theta_{\mathrm{I}} - \frac{\delta}{\varepsilon - 1}.$$

Thus the long-run growth rate (28) applies precisely when it exceeds (27). In other words, the economy grows at the maximal of the two growth rates shown in (27) and (28).

One implication is that prohibiting entry when (26) holds will slow down the growth rate of the economy. Another is that anything that does not improve the growth rate of incumbents by enough to reverse the inequality (26) will have no effect on the growth rate of the economy. It will, of course, have level effects. In a somewhat different economy, this is an important theme of Atkeson and Burstein [2011].

Existence of a Balanced Growth Path To show more precisely that there is in fact a balanced growth path when (26) holds, set $D[S_t/H_t] = 0$ and $D[V_t/C_t] = 0$ in (22) and (23). Combining the resulting equations with (21) gives

$$\frac{q}{w} = \frac{1}{1 - 1/\varepsilon}\frac{L(q/w)}{S/H}\frac{V}{C}, \quad \frac{S}{H} = \frac{E(q/w)}{\eta + \delta - (\varepsilon - 1)(\theta_I - \theta_E)}, \quad \frac{V}{C} = \frac{1/\varepsilon}{\rho + \frac{E(q/w)}{S/H}}.$$

Eliminating S/H and V/C yields

$$\frac{qE(q/w)}{wL(q/w)} = \frac{1}{\varepsilon - 1}\frac{1}{1 + \frac{\rho}{\eta + \delta - (\varepsilon - 1)(\theta_I - \theta_E)}}. \tag{29}$$

The right-hand side is positive because $\rho > 0$, $\varepsilon > 1$ and (26) holds. The left-hand side is strictly increasing, ranging from 0 to ∞ as q/w ranges from 0 to ∞. Hence there will be a unique balanced growth path.

From (25), observe that C_t/H_t is an increasing function of S/H and $L(q/w)$. Thus the level of this balanced growth path is determined population, entry productivity, as well as the equilibrium value of q/w. Note that wages are $w_t = (1-1/\varepsilon)(C_t/H_t)/L(q/w)$, and this is an increasing function of S/H. Use (29) to write

$$\frac{S}{H} = \frac{1}{\varepsilon - 1}\frac{L(q/w)/(q/w)}{\rho + \eta + \delta - (\varepsilon - 1)(\theta_I - \theta_E)}.$$

An increase in θ_I that does not reverse (26) will lower the right-hand side of (29), and hence require a reduction in the equilibrium value of q/w. Since $L(q/w)/(q/w)$ is decreasing in q/w, it follows that S/H and hence wages rise with this type of increase in θ_I. Per-capita consumption increases as well since $L(q/w)$ rises. This describes the level effect of incumbent growth rates alluded to earlier.

Combining the equations for S/H and V/C gives

$$\frac{V}{C} = \frac{1/\varepsilon}{\rho + \eta + \delta - (\varepsilon - 1)(\theta_I - \theta_E)}.$$

This does not depend on the functions $E(\cdot)$ and $L(\cdot)$. The level of aggregate profits is C/ε and growth rates and discount rates do not depend on the relative supplies of entrepreneurial effort and labor.

13

3.7 The Firm Age, Productivity, and Size Distributions

As already noted, initial conditions determined the distribution of productivity when there is no entry. Consider instead the balanced growth path that arises when (26) holds. Recall that $L_t[z]/L_t = (z/Z_t)^{\varepsilon-1}/N_t$, and hence $S_t = N_t(Z_t/X_t)^{\varepsilon-1}$ implies firm employment levels that satisfy

$$\frac{L_t[z]}{L_t/H_t} = \frac{1}{S_t/H_t}\left(\frac{z}{X_t}\right)^{\varepsilon-1}.$$

Away from the balanced growth path, firm employment levels co-move because of fluctuations in L_t/H_t and S_t/H_t. But along the balanced growth path L_t/H_t and S_t/H_t are constant. All firms enter with $z_{t,0}/X_t$ and then firms grow with age according to $z_{t,a}/X_{t+a} = e^{(\theta_I-\theta_E)a}$ as long as productivity does not drop to zero randomly.

The flow of entrants $E_t = Ee^{\eta t}$ implies a population of firms that is growing at the rate η. The random death rate of incumbent firms only affects the size of this population and its age distribution. This age distribution is exponential, with a density $(\eta+\delta)e^{-(\eta+\delta)a}$. At a point in time, the distribution of productivity relative to entry productivity z/X_t is just the distribution of $e^{(\theta_I-\theta_E)a}$. Combining this with the age distribution yields a cross-sectional distribution of productivity relative to entry productivity with the Pareto density

$$f\left(\frac{z}{X_t}\right) = \zeta_*\left(\frac{z}{X_t}\right)^{-(1+\zeta_*)}, \quad \frac{z}{X_t} \geq 1$$

where the tail index ζ_* is given by $\zeta_* = (\eta+\delta)/(\theta_I-\theta_E)$. Since firm revenues and employment are just multiples of $(z/X_t)^{\varepsilon-1}$, this means that the firm size distribution measured using revenues or employment is a Pareto distribution with tail index $\zeta = (\varepsilon-1)\zeta_*$, or

$$\zeta = \frac{\eta+\delta}{(\varepsilon-1)(\theta_I-\theta_E)}. \tag{30}$$

The condition (26) for stationarity corresponds to $\zeta > 1$.

4. THE PHASE DIAGRAM

Assume that the condition (26) for balanced growth holds. To describe precisely how this economy converges to its balanced growth path, it will be helpful to simplify notation. Write $s_t = S_t/H_t$ and $v_t = V_t/C_t$ for the key state variables that appear in the differential

14

equation (21)-(24), and write

$$\gamma = \eta + \delta - (\varepsilon - 1)(\theta_{\mathrm{I}} - \theta_{\mathrm{E}}) = (\eta + \delta)\left(1 - \frac{1}{\zeta}\right) \tag{31}$$

for the key "convergence parameter" that determines whether or not there is a balanced growth path with entry. Here we are assuming that γ is positive, and as noted, this corresponds to $\zeta > 1$. Also define

$$G(y) = E(x), \text{ where } x \text{ solves } \frac{yL(x)/x}{1 - 1/\varepsilon} = 1.$$

From (21), $G(y)$ is the supply of entrepreneurial effort as a function of $y = v_t/s_t$. The function $G(y)$ is strictly increasing, from $G(0) = 0$ to $G(\infty) = E(\infty) \in (0, \infty)$. A fortiori, $yG(y)$ ranges monotonically from 0 to ∞. The elasticity of $G(y)$ is critical for the dynamics of this economy. It can be written as

$$\mathcal{E}(y) = \frac{yDG(y)}{G(y)} = \frac{\frac{xDE(x)}{E(x)}}{1 + \left(-\frac{xDL(x)}{L(x)}\right)}, \text{ where } x \text{ solves } \frac{yL(x)/x}{1 - 1/\varepsilon} = 1.$$

That is, the elasticity of G is the entrepreneurial supply elasticity shrunk towards zero by one plus the labor supply elasticity.

The differential equation (21)-(24) can now be written as

$$\mathrm{D}s_t = -\gamma s_t + G\left(\frac{v_t}{s_t}\right), \tag{32}$$

$$\mathrm{D}v_t = \rho v_t + \frac{v_t}{s_t}G\left(\frac{v_t}{s_t}\right) - \frac{1}{\varepsilon}, \tag{33}$$

together with the transversality condition $e^{-\rho T}v_T \to 0$ as $T \to \infty$. The implied balanced growth path is given by the unique pair (s, v) that solves

$$\gamma s = G\left(\frac{v}{s}\right), \quad v = \frac{1/\varepsilon}{\rho + \gamma}. \tag{34}$$

Observe that $s \uparrow E(\infty)/\gamma$ implies $v \to \infty$ along the curve determined by $\mathrm{D}s_t = 0$. Near the balanced growth path, the dynamical system (32)-(33) can be approximated linearly. Differentiating the right-hand side of (32)-(33) and evaluating the result at the solution to (34) gives

$$\begin{bmatrix} \mathrm{D}s_t \\ \mathrm{D}v_t \end{bmatrix} \approx \begin{bmatrix} -\gamma\left[1 + \mathcal{E}(v/s)\right] & \gamma\mathcal{E}(v/s)/(v/s) \\ -\gamma\left[1 + \mathcal{E}(v/s)\right](v/s) & \rho + \gamma\left[1 + \mathcal{E}(v/s)\right] \end{bmatrix} \begin{bmatrix} s_t - s \\ v_t - v \end{bmatrix}.$$

15

Note the appearance of the factors v/s and $1/(v/s)$ in the off-diagonal elements of the matrix on the right-hand side of this equation. This implies that the eigenvalues of this matrix only depend on v/s via $\mathcal{E}(v/s)$, and not directly. One of the eigenvalues is $-\sigma$, where

$$\sigma = -\frac{\rho}{2} + \sqrt{[1 + \mathcal{E}(v/s)]\left(\gamma + \frac{\rho}{2}\right)^2 - \mathcal{E}(v/s)\left(\frac{\rho}{2}\right)^2}. \tag{35}$$

The fact that ρ and γ are positive ensures that the eigenvalue $-\sigma$ is negative and that the other eigenvalue is positive. Thus there is one stable eigenvalue and one unstable eigenvalue. Convergence to the balanced growth path requires that $[s_t - s, v_t - v]'$ is approximately the eigenvector associated with the stable eigenvalue $-\sigma$, and σ is the approximate exponential rate at which $[s_t - s, v_t - s]$ converges to zero. Thus σ measures the speed of convergence in this economy. Notice that $\sigma = 0$ at $\gamma = 0$ and $\sigma > 0$ whenever $\gamma > 0$. Moreover

$$\left.\frac{\partial \sigma}{\partial \gamma}\right|_{\gamma=0} = 1 + \mathcal{E}(v/s).$$

Thus σ behaves like $[1 + \mathcal{E}(v/s)]\gamma$ for small γ. The elasticity $\mathcal{E}(v/s)$ therefore determines how sensitive the speed of adjustment σ is to the convergence parameter γ. The slope of the stable manifold is given by

$$\left.\frac{\partial v_t}{\partial s_t}\right|_{(s_t,v_t)=(s,v)} = \left(1 + \frac{1}{\mathcal{E}(v/s)}\left[1 - \frac{\sigma}{\gamma}\right]\right)\frac{v}{s}$$

near the balanced growth path. This slope is positive and approaches zero as γ goes to zero. Therefore, if γ is small, then the speed of adjustment will be slow and the stable manifold will be flat.

An example is shown in Figure 2 below. This figure shows the curves $\mathrm{D}s_t = 0$ and $\mathrm{D}v_t = 0$, dashed curves that are not equilibrium trajectories, and a solid curve that represents the equilibrium trajectory for all initial conditions. Observe that $v_t/s_t > v/s$ when $s_t < s$ along the equilibrium trajectory. Since $E(q_t/w_t) = G(v_t/s_t)$, this means that q_t/w_t and entry are high when the economy is below its steady state, as expected. The labor and entrepreneurial supplies in this figure are $L(q/w) = 1/(1 + q/w)$ and $E(q/w) = [(q/w)/(1 + q/w)]^\alpha$, with $\alpha = 2$. The subjective discount rate is $\rho = 0.04$ and $\beta = 0.021$. Given a population growth rate of $\eta = 0.01$ and a random exit rate of $\delta = 0.20$, this corresponds to a tail index $\zeta = 1.11$. The implied speed of convergence is $\sigma = 0.0312$ and the slope of the stable manifold is 0.27. This speed of convergence implies a half-life of 22 years, down from the 33 years that would be implied if the supply of entrepreneurial effort were completely inelastic. Along the balanced growth path, the

16

elasticity $\mathcal{E}(v/s)$ equals $\alpha/(1+q/w) = 1.07$. Doubling α raises this elasticity to 1.7, and the resulting convergence rate rises to $\sigma = 0.0345$, implying a half-life that is still as long as about 20 years.
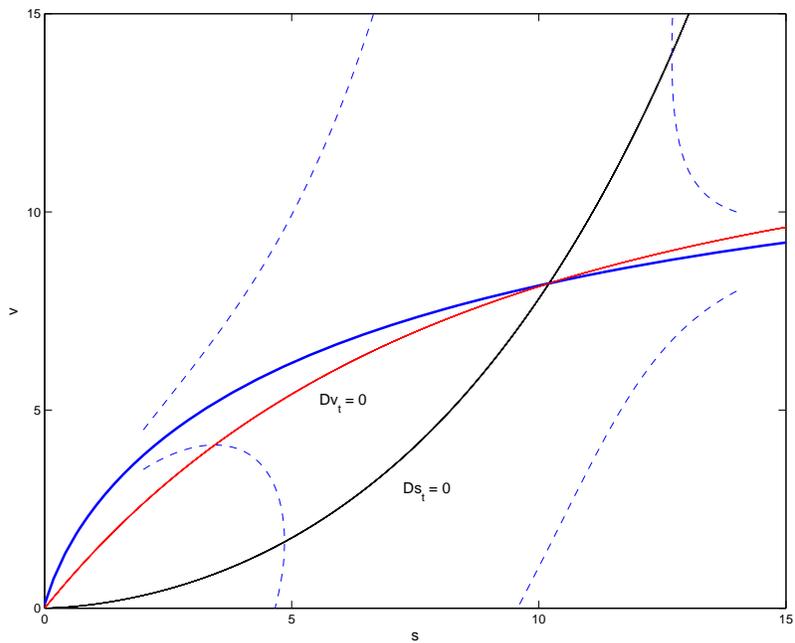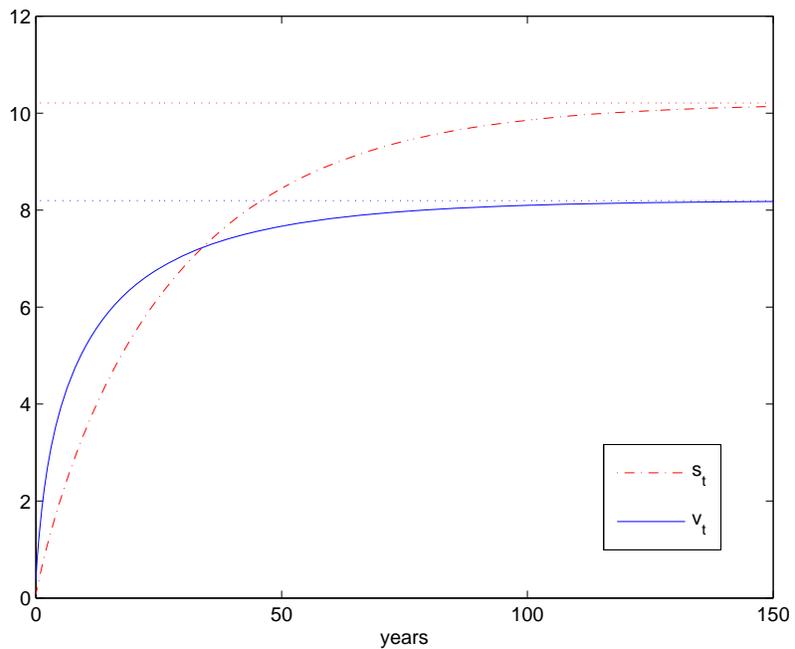

FIGURE 2 The Phase Diagram

FIGURE 3 Rapid Initial Growth

17

The convergence rate shown in (35) is based on a linear approximation near the steady state. To emphasize the local nature of this calculation, Figure 3 shows the paths of s_t and v_t for an initial condition s_0 very close to zero. These paths show that it is possible to go from almost nothing to 80% of the long-run balanced growth path in under 50 years. It then takes the next 50 years to make it up to 95% of the balanced growth path, and another 50 years to close the remaining gap. A model of this type can thus account for both China's rapid growth and the sluggish recoveries that have followed recent recessions in the US.[3] This non-linearity also implies that a rapidly growing economy that suffers a negative shock, say like Korea or Thailand during the Asian financial crisis of 1997-1998, will bounce back relatively quickly, while an advanced economy near its long-run balanced growth path takes a long time to recover.

5. Building up Organization Capital

An alternative interpretation of firm growth is based on capital accumulation rather than productivity growth, as in Luttmer [2011a]. Suppose entrepreneurs create start-up blueprints that enable the production of new goods. Using labor, these blueprints can be replicated, giving rise to additional new differentiated consumption goods. In turn these replicated blueprints can themselves be replicated. The technology for this replication process exhibits constant returns to scale in the number of existing blueprints and labor. Firms start out with an initial start-up blueprint and can grow without bound as blueprints are replicated over time. Firm boundaries are pinned down by an assumption that blueprints cannot be traded.

To focus on convergence mechanisms not already considered, suppose the entrepreneurial supply of new start-up blueprints is completely inelastic. Write E for the per-capita flow of start-up blueprints. To allow for an elastic supply of labor L_t, replace the flow utilities $\ln(C_t/H_t)$ with $\ln(C_t/H_t) - (\xi L_t/H_t)^{1+\beta}/(1+\beta)$, for some $\xi > 0$ and $\beta \geq 0$. The trade-off between consumption and leisure gives rise to the first-order condition

$$\xi \left(\frac{\xi L_t}{H_t}\right)^{\beta} = \frac{w_t}{C_t/H_t}.$$

Thus $1/\beta$ is the Frisch elasticity of the aggregate labor supply. The Euler equation for consumption again gives rise to interest rates $r_t = \rho + \mathrm{D}C_t/C_t$.

---

[3]In response to the first draft of this paper, Andy Atkeson emphasized the need to simultaneously account for rapily growing economies like China.

Labor can be used to produce differentiated products and to create new blueprints. These two uses of labor will be referred to as production labor and managerial labor. Product markets are the same as before, except that productivity is assumed to be constant at some z. When the measure of differentiated consumption goods is N_t, wages are $w_t = (1-1/\varepsilon)zN_t^{1/(\varepsilon-1)}$, as in (5). Write l_t for the amount of production labor used to produce a particular good. From (9), profits per good will be $w_t l_t/(\varepsilon-1)$. Write m_t for the amount of labor used to replicate a particular blueprint. There is one labor market, and market clearing therefore requires that $L_t/N_t = l_t + m_t$. Eliminating L_t from the first-order condition for labor supply and using the fact that aggregate expenditures on production labor must be $w_t N_t l_t = (1 - 1/\varepsilon)C_t$ gives

$$1 = \frac{(l_t + m_t)^\beta l_t}{1 - 1/\varepsilon} \left(\frac{\xi N_t}{H_t}\right)^{1+\beta}. \tag{36}$$

This may be interpreted as an offer curve for production and managerial labor. In the special case of $\beta = 0$, (36) implies $N_t l_t/H_t = (1-1/\varepsilon)/\xi$. If the Frisch elasticity of labor supply is infinite, then per-capita production labor will be constant.

Managerial labor inputs m_t result in new blueprints at the Poisson rate $f(m_t)$, where $f(\cdot)$ is an increasing production function that exhibits decreasing returns to scale. The price of a new blueprint is denoted by q_t, as before. Since all blueprints are the same, this is also the price of every existing blueprint. The aggregate value of all blueprints is thus $V_t = q_t N_t$. The optimal amount of managerial labor used to replicate a given blueprint must satisfy the first-order condition $q_t Df(m_t) = w_t$, assuming the non-negativity constraint on m_t does not bind. Since $w_t N_t l_t = (1 - 1/\varepsilon)C_t$, this can also be written as

$$1 = \frac{V_t}{C_t} \frac{l_t Df(m_t)}{1 - 1/\varepsilon}. \tag{37}$$

As before, V_t/C_t will be a convenient forward-looking state variable. Given the state variables N_t/H_t and V_t/C_t, the static equilibrium conditions (36) and (37) determine the allocation $[l_t, m_t]$ of production and managerial labor. This allocation is uniquely determined: production and managerial labor move in opposite directions along the equilibrium condition (36), and in the same direction along (37).

Note that N_t/H_t only shifts (36) and V_t/C_t only shifts (37). An increase in N_t/H_t must lower both l_t and m_t. Less labor can be used per blueprint when there are more blueprints. Observe that an increase in ξ works in the same way. If consumers put a high value on leisure, then less labor can be used per blueprint. An increase in V_t/C_t raises m_t and lowers l_t. Valuable blueprints cause substitution away from production towards creating new blueprints.

19

Blueprints fail at the rate λ, and so the value of a blueprint must satisfy the asset pricing equation

$$r_t q_t = w_t \left(\frac{l_t}{\varepsilon - 1} - m_t \right) + (f(m_t) - \lambda) q_t + \mathrm{D} q_t.$$

The first term on the right-hand side describes profits from producing the good and the cost of hiring the managers who replicate the blueprint. The second term reflects capital gains that arise when replication succeeds and when the blueprint fails. The third term accounts for capital gains that occur when the state of the economy changes over time.

The state variable N_t/H_t evolves according to

$$\mathrm{D}\left[\frac{N_t}{H_t} \right] = -(\eta + \lambda - f(m_t)) \left[\frac{N_t}{H_t} \right] + E. \tag{38}$$

Combining this with the asset pricing equation for q_t, the Euler condition $r_t = \rho + \mathrm{D}C_t/C_t$, and the fact that $w_t N_t l_t = (1 - 1/\varepsilon) C_t$ gives

$$\mathrm{D}\left[\frac{V_t}{C_t} \right] = \left(\rho + \frac{E}{N_t/H_t} \right) \left[\frac{V_t}{C_t} \right] - \left(\frac{1}{\varepsilon} - \left(1 - \frac{1}{\varepsilon} \right) \frac{m_t}{l_t} \right). \tag{39}$$

The static equilibrium conditions (36) and (37) determine l_t and m_t as functions of the state variables N_t/H_t and V_t/C_t, and then (38) and (39) describe how the state variables evolve over time. An initial condition determines N_0/H_0, and the transversality condition $e^{-\rho t} V_t/C_t \to 0$ for large t provides the other boundary condition.

In contrast to (22)-(23), the mean-reversion coefficient in (38) is now an endogenous variable, and so are the aggregate profits (relative to aggregate consumption) that appear in the last term on the right-hand side of (39).

5.1 Balanced Growth

From (38), a balanced growth path can only arise if

$$\gamma = \eta + \lambda - f(m) \tag{40}$$

is positive at some equilibrium level of managerial labor. Setting the left-hand sides of (38) and (39) equal to zero gives $N/H = E/\gamma$ and $V/C = [(1/\varepsilon) - (1 - 1/\varepsilon) m/l]/(\rho + \gamma)$. Inserting this expression for V/C into the first-order condition (37) and solving for l gives

$$l = (\varepsilon - 1) \left(m + \frac{\rho + \gamma}{\mathrm{D}f(m)} \right). \tag{41}$$

Using this expression for l to eliminate production labor from the labor supply condition (36) yields

$$1 = \left(\frac{m}{\gamma} + \left(1 - \frac{1}{\varepsilon} \right) \frac{1 + \rho/\gamma}{\mathrm{D}f(m)} \right)^{\frac{\beta}{1+\beta}} \left(\frac{m}{\gamma} + \frac{1 + \rho/\gamma}{\mathrm{D}f(m)} \right)^{\frac{1}{1+\beta}} \varepsilon \xi E \qquad (42)$$

Conditions (40) and (42) determine the balanced growth values of m and γ, and then l follows from (41). These equilibrium conditions are only well defined if $\gamma > 0$. In that case, the right-hand side of (42) is increasing in m, and it grows without bound as $f(m)$ approaches $\eta + \lambda$ from below. If $f(0) = 0$, $\mathrm{D}f(0) = \infty$, and $f(m) > \eta + \lambda$ for all large enough m, then there will be a balanced growth path, and it will be unique. Wages and per-capita consumption will grow at the rate $\eta/(\varepsilon - 1)$.

Observe from (42) that the equilibrium conditions for the polar cases of $\beta = 0$ and $\beta = \infty$ differ only by a factor $1/(1 - 1/\varepsilon) > 1$ multiplying the marginal product $\mathrm{D}f(m)$. It follows that m will be higher, and thus γ closer to zero, when the supply of labor is inelastic.

The key comparative static is with respect to ξE. Reducing this parameter, either because labor supply expands, or because entrepreneurial efforts shrink, causes m to increase and γ to decline. Incumbent creation of blueprints increases when labor is abundant and entrepreneurs do not supply many blueprints. Taking $\xi E \downarrow 0$ results in $\gamma \downarrow 0$. Thus the asymptote $\eta + \lambda$ is approached by $f(m)$ from below as entrepreneurial effort becomes small relative to the supply of labor in the economy. In this limiting economy, the allocation of production and managerial labor is pinned down by (40)-(41) together with $\eta + \lambda = f(m)$. As the economy approaches this limit, replication of existing blueprints, as opposed to the introduction of new start-up blueprints, becomes the main source of employment growth. The Frisch elasticity of labor supply no longer has any effect on how much labor is used per blueprint to produce goods and to create new blueprints. Holding fixed ξ and letting E go to zero in (42) gives

$$\lim_{E \downarrow 0} \frac{N}{H} = \lim_{E \downarrow 0} \frac{E}{\gamma} = \frac{1}{\varepsilon \rho \xi} \frac{\mathrm{D}f(m)}{(1 - 1/\varepsilon)^{1/(1+1/\beta)}},$$

where m is determined by $\eta + \lambda = f(m)$. Thus the per-capita stock of blueprints, and hence wages and per-capita consumption, do still depend on the Frisch elasticity $1/\beta$.

5.2 The Firm Size Distribution

Luttmer [2011a] shows that the firm size distribution in this economy has a right tail that behaves like $n^{-\varsigma}$, where n is the number of blueprints or employment per firm, and

$\zeta = \eta/(f(m) - \lambda)$. Thus $\gamma \downarrow 0$ corresponds to $\zeta \downarrow 1$, as before. But here the equilibrium conditions for a balanced growth path ensure that $\gamma > 0$ and $\zeta > 1$, while in the previous sections this depended on an assumption about exogenously given incumbent productivity growth. As will become apparent, the fact that $\gamma > 0$ is forced by equilibrium conditions becomes a dominant factor in determining the convergence rate of this economy when γ is close to zero.

5.3 The Speed of Convergence

The production function f that describes how managerial labor can be used to replicate blueprints is critical for the dynamic properties of this economy. Near the balanced growth path, the two properties of this production function that are important are the factor share and curvature parameters of f,

$$\mathcal{S}(m) = \frac{m \mathrm{D} f(m)}{f(m)} \in (0,1), \quad \mathcal{F}(m) = -\frac{m \mathrm{D}^2 f(m)}{\mathrm{D} f(m)} > 0.$$

The first-order condition $1 = (q_t/w_t) \mathrm{D} f(m_t)$ implies that the elasticities of m_t and $f(m_t)$ with respect to q_t/w_t are equal to $1/\mathcal{F}(m_t)$ and $\mathcal{S}(m_t)/\mathcal{F}(m_t)$, respectively. A sluggish response of blueprint replication arises when f has a lot of curvature, and when the factor share of managerial labor is small. If the elasticity of substitution between managerial labor and blueprints is some constant parameter ω, then $\mathcal{F}(m_t) = (1 - \mathcal{S}(m_t))/\omega$ and hence

$$\frac{\mathcal{S}(m_t)}{\mathcal{F}(m_t)} = \frac{\omega \mathcal{S}(m_t)}{1 - \mathcal{S}(m_t)}.$$

A high managerial labor share then implies a high elasticity of $f(m_t)$ with respect to q_t/w_t, but this will be tempered if the elasticity of substitution ω between managerial labor and blueprints is low.

The more convenient forward-looking state variable is $V_t/C_t = q_t N_t/C_t$ rather than q_t/w_t. As before, write $s_t = N_t/H_t$ and $v_t = V_t/C_t$. Let $m_t = \mathcal{M}[s_t, v_t]$ and $l_t = \mathcal{L}[s_t, v_t]$ denote the solution to the static equilibrium conditions (36)-(37). The elasticities of these equilibrium relations are given by

$$\begin{bmatrix} m_t & 0 \\ 0 & l_t \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{D}\mathcal{M}[s_t, v_t] \\ \mathrm{D}\mathcal{L}[s_t, v_t] \end{bmatrix} \begin{bmatrix} s_t & 0 \\ 0 & v_t \end{bmatrix}$$

$$= -\frac{1}{1 - [1 - \mathcal{F}(m_t)]\left(\frac{(1+\beta)^{-1} + l_t/m_t}{1 + l_t/m_t}\right)} \begin{bmatrix} 1 & -\frac{(1+\beta)^{-1} + l_t/m_t}{1 + l_t/m_t} \\ \mathcal{F}(m_t) & \frac{\beta/(1+\beta)}{1 + l_t/m_t} \end{bmatrix}.$$

22

Thus high curvature of f reduces all elasticities of $\mathcal{M}[s_t, v_t]$ and $\mathcal{L}[s_t, v_t]$, except the elasticity of production labor $\mathcal{L}[s_t, v_t]$ with respect to the per-capita number of blueprints s_t. In the special case of $\beta = 0$, $\mathcal{L}[s_t, v_t]$ only depends on s_t and this elasticity equals -1. The elasticities of $\mathcal{M}[s_t, v_t]$ with respect to s_t and v_t simplify to $-1/\mathcal{F}(m_t)$ and $1/\mathcal{F}(m_t)$, respectively. When the Frisch elasticity of labor supply is infinite, managerial labor responds to the state of the economy as it does to q_t/w_t, and high curvature implies a limited response.

The dynamics of the state variables s_t and v_t are now, from (38) and (39),

$$
\begin{aligned}
\mathrm{D}s_t &= -(\eta + \lambda - f(\mathcal{M}(s_t, v_t)))s_t + E, \\
\mathrm{D}v_t &= \left(\rho + \frac{E}{s_t}\right)v_t - \left(\frac{1}{\varepsilon} - \left(1 - \frac{1}{\varepsilon}\right)\frac{\mathcal{M}(s_t, v_t)}{\mathcal{L}(s_t, v_t)}\right).
\end{aligned}
$$

Near the steady state $[s, v]$ this can be linearized as

$$
\begin{bmatrix} \mathrm{D}s_t \\ \mathrm{D}v_t \end{bmatrix} \approx A \begin{bmatrix} s_t - s \\ v_t - v \end{bmatrix}
$$

where the matrix A is given by

$$
A = \begin{bmatrix} -\gamma & 0 \\ -\gamma v/s & \rho + \gamma \end{bmatrix} + \begin{bmatrix} \mathrm{D}f(m)s & 0 \\ \left(1 - \frac{1}{\varepsilon}\right)l & -\left(1 - \frac{1}{\varepsilon}\right)\frac{m}{l^2} \end{bmatrix} \begin{bmatrix} \mathrm{D}\mathcal{M}[s, v] \\ \mathrm{D}\mathcal{L}[s, v] \end{bmatrix} \tag{43}
$$

and where $m = \mathcal{M}(s, v)$, $l = \mathcal{L}(s, v)$, and $\gamma = \eta + \lambda - f(m)$. Observe that the first matrix on the right-hand side of (43) has eigenvalues $-\gamma$ and $\rho + \gamma$. Absent any movement in the allocation of managerial and production labor per blueprint, this matrix would determine the speed of convergence in this economy. The empirical evidence from the firm size distribution would imply γ close to zero and a very sluggish rate of convergence, even more so than before because the entrepreneurial supply of effort is inelastic.

But here we have to consider the equilibrium responses of $\mathcal{M}[s, v]$ and $\mathcal{L}[s, v]$ as well. Combining (43) with the elasticities of $\mathcal{M}[s, v]$ and $\mathcal{L}[s, v]$ gives

$$
\begin{bmatrix} s & 0 \\ 0 & v \end{bmatrix}^{-1} A \begin{bmatrix} s & 0 \\ 0 & v \end{bmatrix} = \begin{bmatrix} -\gamma & 0 \\ -\gamma & \rho + \gamma \end{bmatrix}
$$
$$
- \frac{\rho + \gamma}{1 - [1 - \mathcal{F}(m)]\frac{(1+\beta)^{-1} + l/m}{1 + l/m}\frac{l/m}{\varepsilon - 1} - 1} \begin{bmatrix} 1 & -\frac{(1+\beta)^{-1} + l/m}{1 + l/m} \\ 1 - \mathcal{F}(m) & -1 \end{bmatrix} \tag{44}
$$

where $\gamma = \eta + \lambda - f(m)$. Observe that this matrix has the same eigenvalues as A. The balanced growth condition (41) can be used to write the ratio l/m that appears in (44)

as

$$\frac{l/m}{\varepsilon - 1} - 1 = \frac{\rho + \gamma}{\eta + \lambda - \gamma} \frac{1}{\mathcal{S}(m)}.$$

With this condition, (44) can be expressed in terms of parameters and the equilibrium values for γ, the factor share $\mathcal{S}(m)$, and the curvature $\mathcal{F}(m)$. The left-hand side of this condition represent net profits per blueprint (production profits minus the cost of replication) scaled by the amount of managerial labor used to replicate blueprints. This profit ratio will be large when the share of managerial labor in replicating blueprints is small.

The speed of convergence implied by (44) is given in the appendix. Here, consider the special case of $\beta = 0$ and take the limit $\gamma \downarrow 0$.[4] Then the matrix (44) simplifies rather drastically to

$$\rho \left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\eta + \lambda}{\rho} \frac{\mathcal{S}(m)}{\mathcal{F}(m)} \begin{bmatrix} 1 & -1 \\ 1 - \mathcal{F}(m) & -1 \end{bmatrix} \right)$$

This matrix has one positive and one negative eigenvalue. This negative eigenvalue implies a speed of convergence equal to

$$\sigma = \frac{\rho}{2} \left(-1 + \sqrt{1 + 4Z} \right), \quad Z = \left(1 + \left(\frac{\eta + \lambda}{\rho} \right) \mathcal{S}(m) \right) \frac{\eta + \lambda}{\rho} \frac{\mathcal{S}(m)}{\mathcal{F}(m)}.$$

This is the formula (2) given in the introduction. An upper bound for σ is $\rho \min\{Z, \sqrt{Z}\}$, and ρZ is the first-order approximation for σ when Z is small. Figure 4 shows the implied half lives for $\rho = 0.02$ and $\eta + \lambda = 0.25$, as a function of the share parameter $\mathcal{S}(m)$ and the curvature parameter $\mathcal{F}(m)$. High blueprints failure rates such as this are used in Luttmer [2011a] to account for the 10% firm exit rate observed in US data. If f is Cobb-Douglas, then $\mathcal{F} = 1 - \mathcal{S}$. This excludes strong curvature in f as a way to account for slow recoveries, although recoveries can still be slow when the factor share of managerial labor is not too high. Figure 4 shows that significantly slower recoveries are possible with curvature parameters above 1. In other words, if the technology for replicating blueprints does not allow for the easy substitution between managerial labor and existing blueprints implicit in the Cobb-Douglas case, then recoveries can be slow.

High curvature is consistent with the absence of any significant co-movement between the state of the aggregate economy and the firm growth rates shown in Figure 1. But

---

[4]The general formula in the appendix indicates that the Frisch elasticity of labor supply has only a limited effect on the convergence rate as long as production labor accounts for a signficant fraction of all labor used per blueprint.

high curvature reduces variation in the managerial inputs used for replication not just over time but also in the cross section. Any model of how firms and the aggregate economy grow will have to account for the cross-sectional variation in firm growth rates as well. The persistent and rapid growth rates shown in Figure 1 are important for aggregate employment growth, but a great many firms do not grow like this. In Luttmer [2011a] such differences in growth rates arise from differences in blueprint quality. With high curvature, this requires large differences in quality. Replication will be easier for some firms than for others, and this type of variation may be needed to account for the cross-sectional heterogeneity in firm growth rates.
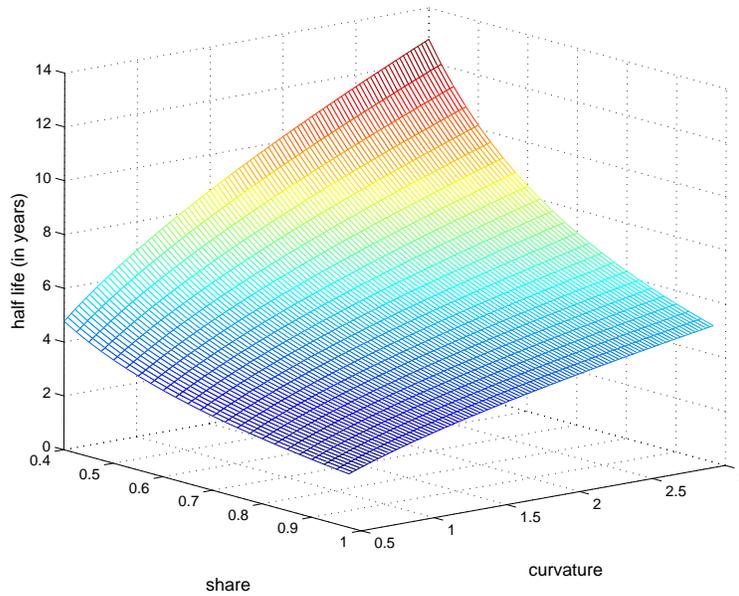

FIGURE 4 Half Lives for $\rho = 0.02$ and $\eta + \lambda = 0.25$.

6. CONCLUDING REMARKS

If the rate of technological progress among incumbent firms does not respond to the state of the economy, then the observed tail index of the firm size distribution implies a very slow rate of convergence unless the elasticity of entry with respect to the value of new firms is large. The contribution of entrants to employment growth in US data is actually remarkably stable, suggesting that shocks that permanently destroy firms or lower incumbent productivity will have long-lasting effects.

25

Slow rates of convergence can also arise in an economy in which firms grow not because of technical change but by accumulating organization capital. An advantage of this alternative interpretation of firm growth is that one can consider shocks that destroy some of the organization capital of incumbent firms, rather than shocks that eliminate entire firms, or hard-to-interpret shocks that lower incumbent productivity. In US data, the negative contribution of firm exit to employment growth is as stable as the positive contribution of entry, and very close in absolute magnitude.

The endogenous response of incumbent rates of organization capital accumulation increases the speed of convergence relative to an economy with exogenous technical change. But Figure 1 tentatively suggests that the scope for this way of speeding up convergence rates is limited. This does raise the question: why do these rapidly growing firms not grow even faster when it is particularly cheap to hire and build up more organization capital? It is hard to believe the firms shown in Figure 1 were held back by limited access to credit or a limited ability to find customers. It is more likely that building up a large organization is a process with costs and benefits that are dominated by long-term considerations.

The labor market in this paper is Walrasian. Introducing search frictions in the labor market, so that firms need to build up not only a stock of projects but also a team of workers matched to these projects, allows one to interpret the relatively slow decline of unemployment that has characterized recent US recessions. This is the subject of ongoing research (Luttmer [2011b]).

A   Convergence with a Finite Frisch Elasticity

If the Frisch elasticity $1/\beta$ is finite and the mean-reversion parameter γ is strictly positive, then the convergence rate implied by (44) is given by

$$\sigma = -\frac{\rho}{2} + \sqrt{\left(\frac{\rho}{2} + \gamma + \frac{(\rho+\gamma)\psi}{1-\phi a}\right)^2 - \frac{(\rho+\gamma)\psi}{1-\phi a}\left(\gamma + \frac{(\rho+\gamma)\psi a}{1-\phi a}\right)\phi}$$

where

$$a = 1 - \mathcal{F}(m) < 1, \quad \phi = \frac{(1+\beta)^{-1} + l/m}{1 + l/m} \in (0,1), \quad \psi = \frac{1}{\frac{l/m}{\varepsilon-1} - 1} > 0.$$

It is not difficult to verify that σ is real and positive. Taking the limit as γ goes to zero and simplifying gives

$$\lim_{\gamma \downarrow 0} \sigma = \frac{\rho}{2} \left(-1 + \sqrt{1 + \frac{4(1+\psi)\psi}{1 - \phi a}} \right).$$

Using the fact that the ratio of profits over managerial inputs $(\varepsilon-1)^{-1}(l/m)-1$ converges to $\rho/[(\eta + \lambda)\mathcal{S}(m)]$ as $\gamma \downarrow 0$, this can be written more explicitly as

$$\lim_{\gamma \downarrow 0} \sigma = \frac{\rho}{2} \left(-1 + \sqrt{1 + 4Z} \right),$$

where

$$Z = \frac{\left(1 + \left(\frac{\eta+\lambda}{\rho} \right) \mathcal{S}(m) \right) \left(\frac{\eta+\lambda}{\rho} \right) \mathcal{S}(m)}{1 + (\mathcal{F}(m) - 1) \left[\frac{(1+\beta)^{-1} + (\varepsilon-1)\left(1 + \left[\left(\frac{\eta+\lambda}{\rho} \right) \mathcal{S}(m) \right]^{-1} \right)}{1 + (\varepsilon-1)\left(1 + \left[\left(\frac{\eta+\lambda}{\rho} \right) \mathcal{S}(m) \right]^{-1} \right)} \right]}.$$

At $\beta = 0$ this simplifies to the case presented in the text. Since m is determined by $\eta + \lambda = f(m)$, one can use this expression to examine how the Frisch elasticity of labor supply affects the convergence rate of this economy. This depends on whether the curvature $\mathcal{F}(m)$ of the production function f is greater or smaller than 1. If $\mathcal{F}(m) > 1$, then Z is increasing in β, and so the highest rate of convergence arises when the aggregate supply of labor is completely inelastic. On the other hand, if $\mathcal{F}(m) < 1$ then Z is decreasing in β and the highest convergence rate arises when the Frisch elasticity is infinite. In any case, the coefficient multiplying $\mathcal{F}(m) - 1$ in the expression for Z varies between $l/(l + m)$ and 1, as β ranges from ∞ to 0. If production labor is an important component of all labor per blueprint, then alternative assumptions about the Frisch elasticity $1/\beta$ will not have an important effect on the convergence rate of this economy.

References

[1] Atkeson, A. and A.T. Burstein, "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, vol. 118, no. 3, 433-484 (2010).

[2] Atkeson, A. and A.T. Burstein, "Aggregate Implications of Innovation Policy," UCLA working paper (2011).

[3] Haltiwanger, J.C., R.S. Jarmin, J. Miranda, "Who Creates Jobs? Small vs. Large vs. Young," *NBER working paper no. 16300* (2010).

[4] Hansen, G.D., "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, vol. 16 (1985), 309-327.

[5] Luttmer, E.G.J., "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, vol. 122, no. 3, 1103-1144 (2007).

[6] Luttmer, E.G.J., "On the Mechanics of Firm Growth," *Review of Economic Studies*, vol. 78, no. 3, 1042-1068 (2011a).

[7] Luttmer, E.G.J., "Firm Growth and Unemployment," Federal Reserve Bank of Chicago seminar, November (2011b).

[8] Moscarini, G. and F. Postel-Vinay, "The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment," *American Economic Review*, forthcoming (2011).

[9] Reinhart, C.M. and K.S. Rogoff, *This Time is Different—Eight Centuries of Financial Folly*, Princeton University Press (2009).

[10] Rogerson, R. "Indivisible Labor, Lotteries and Equilibrium," *Journal of Monetary Economics*, vol. 21 (1988), 3-16.