Slow Convergence in Economies with Organization Capital[1]

please do not circulate: for SED 2017 evaluation only

Erzo G.J. Luttmer
University of Minnesota

February 15, 2017

[1]This is a substantially revised and expanded version of my working paper "Slow Convergence in Economies with Firm Heterogeneity" (Luttmer [2012]).

Abstract

Most firms begin very small and large firms are the result of typically decades of persistent growth. This can be understood as the result of some form of capital accumulation—organization capital. In the US, the distribution of firm size k has a right tail only slightly thinner than $1/k$. This means that most capital accumulation must be accounted for by incumbent firms. This paper describes a range of circumstances in which this implies very slow aggregate convergence rates, in contrast to the rapid convergence of the Cass-Koopmans economy, given standard calibrations of the aggregate labor share. Through the lens of the models described in this paper, the aftermath of the Great Recession of 2008 is unsurprising if the events of late 2008 and early 2009 are interpreted as a destruction of organization capital.

1. Introduction

Most new firms start out with only a few employees, and it has taken the largest firms in the US economy decades of persistent growth to reach their current size. At the same time, the distribution of employment across US firms is very skewed. Although there are as many as 6 million employer firms, about half of aggregate employment is accounted for by the roughly 18,000 firms with more than 500 employees. And the 1,000 or so firms with more than 10,000 employees account for nearly a quarter of aggregate employment. To a first approximation, the distribution of employment size k of US firms is Pareto with a right tail that behaves like $k^{-\zeta}$, with $\zeta \approx 1.10$, just inside the $\zeta > 1$ region where the mean of a Pareto distribution is finite.[1]

These facts are consistent with a very simple model of firm size: there is a constant flow f of new firms that start with size $k = 1$, grow at some rate g, and randomly exit at a mean rate $\varepsilon > g$. An easy calculation, reported in Section 3, shows that this yields $\zeta = \varepsilon/g > 1$. Furthermore, this process of firm entry, growth, and exit implies that the aggregate size K_t evolves according to $\mathrm{d}K_t = -\varepsilon K_t \mathrm{d}t + (gK_t + f)\mathrm{d}t$. That is, the aggregate mean-reversion rate is $\varepsilon - g = (1 - 1/\zeta)\varepsilon$. Holding fixed ε, this implies slow aggregate convergence precisely when $\zeta > 1$ is close to 1, when the size distribution of firms is very skewed. In the US, firm entry and exit rates are around 10% per annum, and so $\varepsilon - g \approx (1 - 1/1.1)0.1 \approx 0.01$. The implied half life is almost 70 years. Even longer half lives can emerge when not all exit is random (Luttmer [2011]).

This model of firm growth provides useful intuition but is too mechanical to be convincing. This paper enriches the model by interpreting firm growth as organization capital accumulation and adding a rich array of margins along which f and g can respond to the state of the economy. The convergence rate of the economy is then characterized in terms factor supply elasticities, factor share parameters, and curvature parameters. This characterization continues to suggest a tight connection between slow aggregate convergence and skewed firm size distributions.

The Cass-Koopmans one-sector model of optimal growth that is the backbone of most of modern macroeconomics does not produce aggregate dynamics of this type and offers no interpretation of the micro evidence on firms. With logarithmic utility and an inelastic labor supply, and following a one-time destruction of capital or a one-time

[1] These facts are well known. See Luttmer [2010] for sources and a survey of some explicit models.

improvement in total factor productivity, it produces a mean reversion rate that is an order of magnitude faster. Combined with a transitory productivity shock it implies a fast recovery that is unlike the relatively slow recoveries observed in the US in recent years. The traditional approach to slow down the rate of convergence of the classic Cass-Koopmans economy is to introduce adjustment costs. But to attribute the small size of many firms to adjustment costs would be akin to saying that most of us are not billionaires because of adjustment costs. Just as dynastic wealth can grow without bound, firms in the economy described here do not have some finite optimal size, but can potentially grow forever.

The model of organization capital described in this paper is, formally, a two-sector economy with joint production: capital can be used simultaneously to produce consumption and new capital. The production of consumption also requires a Leontief composite of managerial services and labor. Replicating organization capital requires not only existing capital but also managerial services. In addition, new capital can be produced from scratch, using a fixed factor (scientists, land, say) together with managerial services. Households supply both labor and managerial services, and these supplies respond to factor prices. Outside the steady state, the wages of managers and workers tend to move in opposite directions if the cost share of managers in a team is substantial. The production from scratch of a unit of start-up capital can be interpreted as firm entry, and a mature firm is then all the organization capital produced from this unit of start-up capital directly or indirectly by replication.

If the supply of labor and managerial services is abundant, then the steady state will have $\zeta > 1$ close to 1, and most of the aggregate accumulation of capital is accounted for by gK_t rather than f. For a given capital stock, both replication and producing capital from scratch are subject to decreasing returns to managerial services. But when labor and managerial services are abundant, the steady state capital stock will be large and offer many opportunities for replicating capital, while the opportunities for entry are limited by the fixed factor. Not surprisingly, the roles of gK_t and f are reversed when the capital stock is far below its steady state.

In this economy, the aggregate factor share of labor and managerial services is a weighted average of the shares in the two sectors of the economy—the consumption sector and the capital producing sector. If these labor and managerial services shares are the same, and if $\zeta > 1$ is very close to 1, then the aggregate convergence rate of this economy is not 0, as it would be if f and g are constants, but about half the convergence rate of a standard one-sector economy with the same labor share. But the convergence rate in the $\zeta \approx 1$ version of the two-sector economy is an increasing function of the factor share of managerial services in replicating capital. In particular, the aggregate convergence rate is close to zero when the managerial factor share is small, as in the case of inelastic f and g. The convergence rate of a standard one-sector economy also vanishes as the factor share of labor goes to zero. This gives rise to the familiar AK model of long-run growth. But labor shares in advanced economies are high, traditionally calibrated at 70%. Recent declines in these labor shares are not enough to significantly affect the convergence rate of the one-sector economy. In the economy with organization capital, however, a large labor share in the consumption-producing sector can be combined with a low managerial services share in the sector replicating capital, resulting in a slow aggregate convergence rate.

The aggregate convergence rate of the $\zeta \approx 1$ version of the organization capital economy not only depends on the factor share of managerial services but also on the curvature of the replication rate g as a function of managerial services. High curvature will make g respond inelastically to the aggregate state and lead to slow convergence even if f has very little curvature. When $\zeta \approx 1$, it is a matter of accounting that incumbent replication must make up the bulk of the production of new capital and, and to an extent that is perhaps not widely appreciated, entry can do very little. A particularly elastic residual supply of managerial services (services not used to supervise labor) interacts with high curvature to slow down the economy even further.

Related Literature The connection between the thickness of the right tail of the firm size distribution and the aggregate convergence rate of an economy was first made in Luttmer [2012]. Gabaix et al. [2016] use the same idea to explain why the simplest random growth model cannot account for the fairly rapid changes observed in the US earnings and wealth distributions. The current paper remains focused on firms and expands the number of margins along which the economy can respond to a low stock of organizational capital: the supplies of labor and managerial services depend on factor prices, managers can switch between tasks, and both entry and the rate at which incumbent firms grow are determined endogenously. This leads to a richer understanding of what can give rise to a slow aggregate convergence rate.

Outline Section 2 describes the economy. Section 3 describes the firm size implications. The aggregate equilibrium conditions and the steady state are in Section 4. The characterization of the speed of convergence is in Section 5. The quantitative properties of the economy are explored further in Section 6 and extensions that are crucial for a full empirical implementation are sketched in Section 7.

2. The Economy

The technologies for producing consumption and new capital are different. Households supply labor and managerial services. Labor is a specific factor for the consumption sector, while managerial services can be reallocated across the two sectors.

2.1 Households

Consider an economy with a unit continuum of identical infinitely lived households who consume and supply primary factors of production. Each household is made up of a heterogeneous continuum of members with types indexed by $h = (h_c, h_u, h_v, h_w) \in \mathbb{R}^4_{++}$. The distribution of types in each household is Ψ, a continuous distribution. The components h_c and h_u of the household member type are utility weights. A type-h household member who participates in the labor market can supply either h_v units of managerial services or h_w units of labor. It will be convenient to normalize the mean of h_c to be equal to 1. Household preferences over type-specific consumption flows $C_t(h) \geq 0$ and labor market participation decisions $\iota_t(h) \in \{0, 1\}$ are determined by

$$\mathcal{U}(C, \iota) = \int_0^\infty e^{-\rho t} \left(\int [h_c \ln(C_t(h)) + h_u(1 - \iota_t(h))] \, \Psi(\mathrm{d}h) \right) \mathrm{d}t.$$

Households can earn $\widetilde{w}_t$ per unit of labor and $\widetilde{v}_t$ per unit of managerial services, both measured in units of consumption per unit of time. Households are endowed with an equal share of the assets in this economy, and markets are complete. As a result, every household will consume the same amount of consumption C_t and supply the same amounts of labor and managerial services. But the consumption and factor supplies of individual household members are heterogeneous, as long as Ψ is not degenerate.

The risk-free rate in this economy is related to aggregate consumption growth via the usual Euler condition $r_t = \rho + \mathrm{D}C_t/C_t$.

2.1.1 Factor Supply Curves

Let λ_t denote the marginal utility of wealth of the typical household at time t. Consumption of the individual household members will have to satisfy the first-order condition $h_c = \lambda_t C_t(h)$. Because the mean of h_c is normalize to 1, this implies $1 = \lambda_t C_t$, where C_t is aggregate (and household) consumption. The potential earnings of a type-h household member are $\max\{\widetilde{v}_t h_v, \widetilde{w}_t h_w\}$, and it is optimal for this household member to participate in the labor market if and only if $\lambda_t \max\{\widetilde{v}_t h_v, \widetilde{w}_t h_w\} \geq h_u$. The smooth heterogeneity assumed here means that ties do not affect aggregate factor supplies and avoids the need to consider the employment lotteries proposed in Rogerson [1988]. Since $\lambda_t = 1/C_t$ this can also be written as $\max\{v_t h_v, w_t h_w\} \geq h_u$, where

$$(v_t, w_t) = (\widetilde{v}_t, \widetilde{w}_t)/C_t$$

is the vector of marginal-utility weighted factor prices. Throughout the rest of the paper, "factor prices" or "wages" will always refer to these marginal-utility weighted prices. The resulting aggregate supplies of labor and managerial services are then

$$L(v_t, w_t) = \int h_w \iota \left[w_t h_w > \max\{h_u, v_t h_v\} \right] \Psi(\mathrm{d}h) \tag{1}$$

$$M(v_t, w_t) = \int h_v \iota \left[v_t h_v > \max\{h_u, w_t h_w\} \right] \Psi(\mathrm{d}h) \tag{2}$$

Observe that these supply functions only depend on the marginal-utility weighted factor prices, and nothing else. This relies heavily on the assumption that households are identical. Without such an assumption, aggregate factor supplies would depend on the equilibrium distribution of wealth across households. Taking for granted that the distribution Ψ is nice, it is easy to see that an increase in w will increase $L(v, w)$ and lower $M(v, w)$, and the reverse for an increase in w. Moreover, a proportional increase in both (v, w) will increase both the supply of labor and the supply of managerial services. Precise conditions are given in the following lemma.

Assumption 1 *The type distribution Ψ is continuous, with a finite mean and a strictly positive density.*

Lemma 1 *Suppose Assumption 1 holds. Then* $L(v,0) = M(0,w) = 0$ *for all positive* v *and* w*. Furthermore, both* $L(v,w)$ *and* $M(v,w)$ *are differentiable and the slopes of these supply curves satisfy*

$$\begin{bmatrix} \mathrm{D}M(v,w) \\ \mathrm{D}L(v,w) \end{bmatrix} = \begin{bmatrix} + & - \\ - & + \end{bmatrix}$$

and this matrix is symmetric. In addition $L(\theta v, \theta w)$ *and* $M(\theta v, \theta w)$ *are both strictly increasing in* $\theta > 0$*. This implies that own price elasticities are larger in absolute value than cross price elasticities.*

To simplify notation, write $[\mathcal{E}_{L,v,t}, \mathcal{E}_{L,w,t}]$ and $[\mathcal{E}_{M,v,t}, \mathcal{E}_{M,w,t}]$, respectively, for the elasticities of the labor and managerial services supply functions, evaluated at equilibrium prices. The last observation in Lemma 1 then says that $-\mathcal{E}_{L,v,t}/\mathcal{E}_{L,w,t}$ and $-\mathcal{E}_{M,w,t}/\mathcal{E}_{M,v,t}$ are both in $(0,1)$.

The factor supplies (1)-(2) are the result of households solving a discrete-choice problem. Everyone who works in this economy "works hard," and the aggregate factor supply elasticities are completely driven by the numbers of households who are at the margins between not working, supplying labor, and supplying managerial services. The only household members who move in and out of the labor force with fluctuations in the state of the economy are those who have both low h_v/h_u and h_w/h_u. For household members with high h_v/h_u and h_w/h_u, the important margin is whether to supply labor or managerial services. Household members can switch back and forth over time depending on the relative price $v_t/w_t = \widetilde{v}_t/\widetilde{w}_t$. It is easy to imagine how this type of substitution across occupations can happen within a given employment relationship. This poses an important challenge for measurement.

2.2 The Technology for Producing Consumption

Consumption is produced using capital K_t and the services of a team of managers and workers. Team services are produced using a Leontief technology that combines β units of managerial services with every unit of labor services. In any equilibrium, managerial and labor services will be used in exactly this proportion, and so $L_t = L(v_t, w_t)$ measures both labor and team services. The output of consumption is then $C_t = F(K_t, L_t)$, where F is a production function that is assumed to be strictly increasing in both factors, smooth, and that exhibits constant returns to scale. Because of the logarithmic utility assumption, the production function F will turn out to affect the dynamic properties of this economy only via the function

$$A(k) = \frac{\mathrm{D}_2 F(k,1)}{F(k,1)}.$$

This is the factor share of a team services when $K_t/L_t = k$. The Leontief technology for team services implies that the marginal-utility weighted cost of a team is $\beta v_t + w_t$ per unit of team services. Equating the cost of a team with its marginal product and clearing the labor market gives

$$(\beta v_t + w_t) L\left(v_t, w_t\right) = A\left(\frac{K_t}{L(v_t, w_t)}\right). \tag{3}$$

This can be viewed as a condition that determines a function $w_t = W(K_t, v_t)$ that describes how wages vary with K_t and v_t in any equilibrium.[2] It is easy to see that a destruction of capital has a negative direct effect (that is, holding fixed v_t) on w_t and $L(v_t, w_t)$ if $A(\cdot)$ is increasing. This will be the maintained assumption.

Assumption 2 *The production function F for consumption is strictly increasing in capital and team services, sufficiently smooth, concave, and exhibits constant returns to scale. The implied factor share of team services $A(\cdot) \in (0,1)$ is non-decreasing.*

The team factor share is decreasing in K_t/L_t if F is a CES production function with an elasticity strictly below 1. The function $A(\cdot)$ is constant if F is Cobb-Douglas, and then the dependence of the right-hand side of (3) on K_t vanishes. The capital stock can still affect wages and worker employment, but only indirectly through the price v_t of managerial services. If $\beta > 0$ then there are two such indirect channels: an increase in v_t raises the cost of a team of managers and workers, and, holding fixed w_t, it lowers the supply of labor because marginal households switch from supplying labor to supplying managerial services. If $\beta = 0$ then (3) implies that v_t and w_t co-move, weakly. If the cross-price elasticities of $L(v_t, w_t)$ and $M(v_t, w_t)$ are also zero, then $w_t = \widetilde{w}_t/C_t$ and the

[2] What is very special here is that F depends on labor and managerial services only through the team services composite good. The Leontief assumption can be relaxed. If the technology for team services has a unit cost function $T(v, w)$, then (3) generalizes to $T(v_t, w_t)L(v_t, w_t)/\mathrm{D}_2T(v_t, w_t) = A(K_t/[L(v_t, w_t)/\mathrm{D}_2T(v_t, w_t)])$.

supply of labor are both constant. It will be interesting not to impose these rather stark simplifying assumptions a priori but examine how these elasticities affect the response of the economy to a destruction of capital.

Lemma 2 *Write $\mathcal{E}_{A,t}$ for the elasticity of the labor share function $A(k_t)$, and $\mathcal{E}_{W,K,t}$ and $\mathcal{E}_{W,v,t}$ for the elasticities of the wage function $w_t = W(K_t, v_t)$ defined by* (3). *Assumptions 1 and 2 imply that $W(K_t, v_t)$ is well defined and*

$$\left[\begin{array}{cc} \mathcal{E}_{W,K,t}, & \mathcal{E}_{W,v,t} \end{array}\right] = \frac{\left[\begin{array}{cc} \mathcal{E}_{A,t}, & -\left(\frac{\beta v_t}{\beta v_t + w_t} + (1+\mathcal{E}_{A,t})\,\mathcal{E}_{L,v,t}\right) \end{array}\right]}{\frac{w_t}{\beta v_t + w_t} + (1+\mathcal{E}_{A,t})\,\mathcal{E}_{L,w,t}}.$$

Although the sign of $\mathcal{E}_{W,v,t}$ is ambiguous, the capital-labor ratio $k_t = K_t/L(v_t, W(K_t, v_t))$ is increasing in v_t holding fixed K_t, and increasing in K_t along any increasing curve $K_t \mapsto v_t$.

At a point in time, holding fixed the capital stock K_t, Lemma 2 implies that the factor prices v_t and w_t move in opposite directions if $\beta > 0$ and there are relatively few households on the margin between supplying labor and managerial services.

Using the wage function $w_t = W(K_t, v_t)$ implied by (3), define

$$S(K_t, v_t) = M(v_t, W(K_t, v_t)) - \beta L(v_t, W(K_t, v_t)). \tag{4}$$

This is the residual supply of managerial services available to the sector producing new capital. It is immediate that $S(K_t, v_t)$ is weakly decreasing in K_t under Assumptions 1 and 2. An increase in the capital stock raises the wages of workers, by Lemma 2, and this unambiguously reduces the supply of managerial services and raises the demand for managerial services that arises from team production. It is also immediate that $S(K_t, v_t)$ is increasing in v_t if $w_t = W(K_t, v_t)$ is decreasing in v_t. It can be shown that this result remains true if w_t and v_t co-move, holding fixed K_t.

Lemma 3 *Assumptions 1 and 2 imply that the residual supply of managerial services $S(K_t, v_t)$ is weakly decreasing in K_t (strictly if and only if $A(\cdot)$ is strictly increasing) and strictly increasing in v_t.*

The proof is in Appendix A. The residual supply of managerial services can be constant if we relax Assumption 1 and allow $M(\cdot, \cdot)$ to be constant. This implies that $L(v_t, w_t)$

cannot depend on v_t, and then a constant residual supply of managerial services will arise if $\beta = 0$ or if $A(\cdot)$ is constant. It will be convenient to write

$$\left[\ \mathcal{E}_{S,K,t},\quad \mathcal{E}_{S,v,t}\ \right] = \left[\ \frac{\mathrm{D}_1 S(K_t,v_t)K_t}{S(K_t,v_t)},\quad \frac{\mathrm{D}_2 S(K_t,v_t)v_t}{S(K_t,v_t)}\ \right]$$

for the elasticities of managerial services with respect to K_t and v_t, at their equilibrium values.

2.3 The Technology for Producing New Capital

There are two ways to produce new capital. Managerial services can be used to produce new capital from scratch. Using n_t units of managerial services generates an aggregate flow of $f(n_t)$ of new units of capital. The production function f is subject to decreasing returns to scale. For example, it could be that managerial services have to be combined with a production location, and that these production locations are heterogeneous and in fixed supply.

The second way capital can be produced is by replicating existing capital. One unit of capital can be replicated using m_t units of managerial services, at the average rate $g(m_t)$. Capital is assumed to be homogeneous, and so the aggregate output of new capital produced by replication is $K_t g(m_t)$. The production function g exhibits decreasing returns to scale. The aggregate stock of capital then evolves according to

$$\mathrm{D}K_t = (g(m_t) - \delta)K_t + f(n_t). \tag{5}$$

The production functions f and g are taken to satisfy the following assumption.

Assumption 3 *The production functions f and g are strictly increasing, strictly concave, and smooth. Furthermore,*

(i) $f(0) = g(0) = 0$,

(ii) *the marginal products* $\mathrm{D}f(n)$ *and* $\mathrm{D}g(m)$ *range throughout* $(0, \infty)$.

Part (i) of this assumption implies that managerial services are essential inputs in producing capital. In particular, $g(0) = 0$ means that the type of autonomous growth of capital that occurs in the AK economies of Jones and Manuelli [1990] and Rebelo [1990] cannot happen here. Part (ii) serves to rule out corner solutions—there will always be some entry and some replication.

2.3.1 The Price of Capital

There is joint production in this economy: the same unit of capital is combined, simultaneously, with labor to produce consumption, and with managerial services to replicate capital. Both activities generate income that accrues to the owners of capital. Write $\widetilde{q}_t$ for the price of a unit of capital, measured in units of consumption. The usual asset pricing equation says that

$$r_t\widetilde{q}_t = \frac{(1-A(k_t))C_t}{K_t} + \max_m \{\widetilde{q}_t(g(m)-\delta) - \widetilde{v}_t m\} + \mathrm{D}\widetilde{q}_t,$$

where $k_t = K_t/L(v_t, W(K_t, v_t))$ is the capital-labor ratio in the consumption sector. That is, the required return on a unit of capital comes in the form of earnings from producing consumption, of earnings from replicating capital, and in the form of capital gains. The first-order condition for replicating capital is $\widetilde{v}_t = \widetilde{q}_t\mathrm{D}g(m_t)$, and the first-order condition for producing capital from scratch is $\widetilde{v}_t = \widetilde{q}_t\mathrm{D}f(n_t)$. Write $q_t = \widetilde{q}_t/C_t$ for the marginal-utility weighted price of a unit of capital. Using the Euler condition $r_t = \rho + \mathrm{D}C_t/C_t$ then gives

$$\rho q_t = \frac{1-A(k_t)}{K_t} + \max_m \{q_t(g(m)-\delta) - v_t m\} + \mathrm{D}q_t.$$

The first-order conditions for producing capital are

$$q_t\mathrm{D}f(n_t) = v_t = q_t\mathrm{D}g(m_t).$$

This is where Assumption 3 comes in to avoid the need to consider corners. As is standard, the optimality conditions for this economy also include a transversality condition that requires $e^{-\rho t}q_tK_t$ to go to zero as t becomes large.

The following shorthand will be useful

$$\mathcal{S}_{g,t} = \frac{\mathrm{D}g(m_t)m_t}{g(m_t)}, \quad \mathcal{S}_{f,t} = \frac{\mathrm{D}f(n_t)n_t}{f(n_t)}, \quad \mathcal{C}_{g,t} = -\frac{\mathrm{D}^2g(m_t)m_t}{\mathrm{D}g(m_t)}, \quad \mathcal{C}_{f,t} = -\frac{\mathrm{D}^2f(n_t)n_t}{\mathrm{D}f(n_t)}.$$

These are the factor share and curvature parameters of the production functions $g(\cdot)$ and $f(\cdot)$, evaluated at equilibrium values of the inputs m_t and n_t. Cobb-Douglas production functions are only the simplest production functions that satisfy Assumption 3. But they are quite restrictive: their share and curvature parameters are constant and have to add up to 1. This automatically rules out the possibility of sharply decreasing returns to scale. Assumption 3 is also satisfied if the production function is Cobb-Douglas up to some threshold and has a constant elasticity of substitution less than 1 for all large managerial inputs above that threshold. Such a production function can have curvature greater than 1 and will have a vanishingly small managerial factor share for large enough managerial inputs.

3. Aggregate Convergence and the Firm Size Distribution

This economy will be shown to have a unique steady state, with $(m_t, n_t) = (m, n)$, and both $f(n)$ and $\delta - g(m)$ positive. That is, existing capital is replicated at a lower rate than the depreciation rate δ, and capital produced from scratch makes up the difference.

The flow $f(n)$ of new capital produced from scratch can be interpreted as a flow of new firms, each with one unit of start-up capital. Firms then grow by replicating capital. Because the allocation of capital across firms does not matter, an arbitrarily small transaction cost is enough to keep all capital produced directly or indirectly from the initial unit of start-up capital within in the same firm. From (5), note that $K = f(n)/(\delta - g(m))$ in the steady state. The contribution of firm entry to aggregate investment is thus

$$\frac{f(n)}{g(m)K + f(n)} = 1 - \frac{g(m)}{\delta} \in (0, 1).$$

So the contribution of entry will be small precisely when $\delta - g(m) > 0$ is close to zero. Of course, even though new firms contribute very little to aggregate capital accumulation, their subsequent contribution as incumbents can be very large.

Suppose now that the capital embodied in firms can depreciate in two distinct ways: incumbent firm capital depreciates continuously at a rate $\delta_k \in [0, \delta)$, and all of the firm's capital is destroyed simultaneously and randomly at the complementary rate $\delta_f = \delta - \delta_k \in (0, \delta)$. That is, δ_f is a firm exit rate, and a firm's exit goes together with the destruction of all of its capital. In a steady state, this means that the age distribution of firms is exponential with mean $1/\delta_f$. Incumbent firms grow at the net rate $g(m) - \delta_k$ as long as the random exit shock does not hit, and so the size of a firm of age a will be $k = e^{(g(m)-\delta_k)a}$, measured in units of capital. Assume that $g(m) - \delta_k$ is positive, so that firms can grow beyond their start-up size. The distribution Φ of firm size will then be

$$\Phi(k) = 1 - e^{-\delta_f \ln(k)/(g(m)-\delta_k)} = 1 - k^{-\zeta}, \quad k \in [1, \infty).$$

This is a Pareto distribution on $[1, \infty)$ and

$$\zeta = \frac{\delta_f}{g(m) - \delta_k}$$

is the tail index of the distribution. The mean of this distribution is finite if and only if $\zeta > 1$. The limiting distribution associated with $\zeta = 1$ is known as Zipf's law. The steady state implies $0 < \delta - g(m) = \delta_f - (g(m) - \delta_k)$, and the assumption is that the equilibrium is such that $g(m) - \delta_k > 0$, so that incumbent firms do in fact grow. Together, this implies that $\zeta > 1$. Moreover, as long as $g(m) - \delta_k > 0$ is bounded away from zero, $\delta - g(m) \downarrow 0$ is the same as $\zeta \downarrow 1$.

In this economy, the capital of incumbent firms depreciates continuously at the rate δ_k. If, instead, individual pieces of capital of incumbent firms depreciate randomly in one-hoss-shay fashion at the rate δ_k, the resulting firm size distribution will be not be Pareto but an analog of the Pareto distribution that has discrete support (a generalization of the Yule distribution associated with the special case $\delta_k = 0$.) In particular, the right tail of that distribution will still behave like $k^{-\zeta}$ (Luttmer [2011]).

Firm employment scales with firm capital because, in the steady state, capital-labor ratios are constant in both the production of consumption and in the replication of capital. In US data, the employment size distribution of firms has a tail index ζ of about 1.06 (Luttmer [2007]) and the interpretation given here means that $\delta - g(m)$ must be small. From (5), K_t converges at precisely the rate $\delta - g(m)$ when (m_t, n_t) is fixed at the steady state value (m, n). That is, slow aggregate convergence happens precisely when the tail index ζ of the firm size distribution is close to 1, as is the case in US data.

This simple calculation needs to be amended to take into account the fact that (m_t, n_t) is likely to depend on the state of the economy. This will speed up the aggregate rate of convergence. This then raises two questions: what features of the economy will result in a small steady state value of $\delta - g(m)$, and what features will result in slow aggregate convergence even if (m_t, n_t) depends on the state of the economy?

4. Equilibrium Conditions and the Steady State

The equilibrium trajectories of this economy are determined by a pair of differential equations that describe the dynamics of K_t and q_t. These differential equations are, from (5) and the asset pricing equation for q_t,

$$\begin{aligned} \mathrm{D}K_t &= -(\delta - g(m_t))K_t + f(n_t), && (6) \\ \mathrm{D}q_t &= (\rho + \delta - (g(m_t) - \mathrm{D}g(m_t)m_t))\, q_t - \frac{1 - A(k_t)}{K_t}, && (7) \end{aligned}$$

where (m_t, n_t, v_t) is determined by the static equilibrium conditions,

$$1 = \frac{\mathrm{D}f(n_t)}{\mathrm{D}g(m_t)}, \quad m_t K_t + n_t = S(K_t, v_t), \quad v_t = q_t \mathrm{D}g(m_t). \tag{8}$$

and k_t is defined as

$$k_t = \frac{K_t}{L\,(v_t, W(K_t, v_t))}. \tag{9}$$

A transversality condition requires that $e^{-\rho t} q_t K_t \to 0$ as t becomes large. The marginal utility weighted wages $w_t = W(K_t, v_t)$ follow from the labor market clearing condition (3) and aggregate consumption is $C_t = F(K_t, L(v_t, W(K_t, v_t)))$.

The derivatives $\mathrm{D}K_t$ and $\mathrm{D}q_t$ given in (6)-(7) are explicit functions of the state (K_t, q_t), the allocation of managerial services (m_t, n_t), and the capital-labor ratio k_t. The allocation of (m_t, n_t) follows from the static equilibrium conditions (8), and then (9) determines k_t. This pins down the trajectory of (K_t, q_t) given an initial value (K_0, q_0). The initial value of K_0 is given, and, as usual, the transversality condition will be needed to determine q_0.

Observe that (6)-(9) depends on the production function F only via its labor share $A(k_t)$, directly by how it affects the earnings of a unit of capital, in (7), and indirectly via $S(K_t, v_t)$, in the static equilibrium conditions (8). In a steady state, k_t is constant and so is $A(k_t)$. As in the one-sector Cass-Koopmans economy, the steady state of this economy can tell us nothing about the shape of $A(\cdot)$.

4.1 The Static Equilibrium Conditions

Since the residual supply curve $S(K_t, \cdot)$ is strictly increasing, it will have an inverse supply curve $S^{-1}(\cdot)[K_t]$. The static equilibrium conditions (8) are the first-order conditions for the surplus maximization

$$\max_{m,n} \left\{ q_t \left(g(m) K_t + f(n) \right) - \int_0^{mK_t + n} S^{-1}(s)[K_t] \mathrm{d}s \right\},$$

provided that this maximization problem has an interior solution. The fact that the supply of managerial services is increasing in v_t implies that this is a concave programming problem with solutions that can be characterized by first-order conditions. The specifics of Assumption 3 imply that the solution is unique and interior.

To see this, note that part (ii) of Assumption 3 implies that the inverse marginal product functions $[\mathrm{D}f]^{-1}$ and $[\mathrm{D}g]^{-1}$ are well defined. One way to summarize the static equilibrium conditions (8) is therefore $[\mathrm{D}g]^{-1}(u_t)K_t + [\mathrm{D}f]^{-1}(u_t) = S(K_t, q_t u_t)$, together with $u_t = \mathrm{D}g(m_t) = \mathrm{D}f(n_t)$. The supply of managerial services $S(K_t, q_t u_t)$ is strictly increasing in u_t, with a range that satisfies $0 \in [S(K_t, 0), S(K_t, \infty))$. Both $[\mathrm{D}g]^{-1}(u_t)$ and $[\mathrm{D}f]^{-1}(u_t)$ are strictly decreasing, and these functions range throughout $(0, \infty)$. It follows that the equation for u_t will have a unique solution given any positive (K_t, q_t). The solution for m_t and n_t then follows. The differential equation (6)-(7) is therefore well defined for all $(K_t, q_t) \in \mathbb{R}^2_{++}$.

Inspection of the static equilibrium conditions (8) shows that m_t and n_t must co-move. The first-order conditions $v_t = q_t \mathrm{D}g(m_t)$ and $v_t = q_t \mathrm{D}f(n_t)$ imply

$$\begin{bmatrix} \frac{K_t}{m_t}\frac{\partial m_t}{\partial K_t} & \frac{q_t}{m_t}\frac{\partial m_t}{\partial q_t} \\ \frac{K_t}{n_t}\frac{\partial n_t}{\partial K_t} & \frac{q_t}{n_t}\frac{\partial n_t}{\partial q_t} \end{bmatrix} = \begin{bmatrix} 1/\mathcal{C}_{g,t} \\ 1/\mathcal{C}_{f,t} \end{bmatrix} \begin{bmatrix} -\frac{K_t}{v_t}\frac{\partial v_t}{\partial K_t} & 1 - \frac{q_t}{v_t}\frac{\partial v_t}{\partial q_t} \end{bmatrix}, \tag{10}$$

where the elasticities on the right-hand side are the elasticities of the managerial factor price v_t that solves the static equilibrium conditions (8) in terms of (K_t, q_t). The matrix (10) is singular because the efficiency condition $1 = \mathrm{D}f(n_t)/\mathrm{D}g(m_t)$ does not depend on (K_t, q_t). This same efficiency condition implies that the relative magnitude of the elasticities of m_t and n_t depends on the relative curvatures of $g(\cdot)$ and $f(\cdot)$: m_t responds less than n_t if the curvature of $g(\cdot)$ exceeds that of $f(\cdot)$, and vice versa.

An explicit calculation of the elasticities of v_t with respect to (K_t, q_t) gives

$$\begin{bmatrix} \frac{K_t}{v_t}\frac{\partial v_t}{\partial K_t}, & \frac{q_t}{v_t}\frac{\partial v_t}{\partial q_t} \end{bmatrix} = \frac{\begin{bmatrix} \frac{m_t K_t}{m_t K_t + n_t} - \mathcal{E}_{S,K,t}, & \frac{m_t K_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{g,t}} + \frac{n_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{f,t}} \end{bmatrix}}{\mathcal{E}_{S,v,t} + \frac{m_t K_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{g,t}} + \frac{n_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{f,t}}}. \tag{11}$$

Recall that $\mathcal{E}_{S,K,t} \leq 0$ and $\mathcal{E}_{S,v,t} > 0$ under Assumptions 1 and 2. So the elasticity of v_t with respect to K_t is positive, and the elasticity of v_t with respect to q_t is in $(0, 1)$. The intuition is that an increase in q_t is distributed over an increase in the price $v_t = q_t \mathrm{D}g(m_t)$ and an increase in the quantity m_t. In turn, this is a consequence of the fact that m_t and n_t must co-move, together with the fact that the supply of managerial services is upward sloping.

Combining (10) and (11) shows that m_t and n_t are both decreasing in K_t and increasing in v_t. As expected from $1 = \mathrm{D}f(n_t)/\mathrm{D}g(m_t)$, $\mathcal{C}_{f,t} \downarrow 0$ drives the elasticities of m_t with respect to K_t and q_t to zero. Neither of the inputs (m_t, n_t) responds much to changes in q_t if the residual supply of managerial services is particularly inelastic. But even with an inelastic residual supply of managerial services, something has to give when K_t changes, and the response of m_t to K_t becomes unit elastic (with a negative sign) when $m_t K_t/(m_t K_t + n_t)$ is close to 1 and $\mathcal{E}_{S,K,t}$ is small (as in the Cobb-Douglas case.)

The elasticities (10) and (11) can be used to prove the following lemma, which will be used below to study the phase diagram of this economy.

Lemma 4 *Suppose Assumptions 1-3 hold and consider* $(\mathrm{D}K_t, \mathrm{D}q_t)$ *as a function of* (K_t, q_t)*. The elasticities (10) and (11) imply that* $\rho = \partial \mathrm{D}K_t/\partial K_t + \partial \mathrm{D}q_t/\partial q_t$*, and that* $\partial \mathrm{D}K_t/\partial q_t > 0$ *and* $\partial \mathrm{D}q_t/\partial K_t > 0$.

The first of these results holds in any economy with additively separable preferences (for example, differentiate the modified Hamiltonian described in Brock and Scheinkman [1976, p.165]). A brute force calculation can be used to verify that it does indeed hold here, taking into account that, by Lemma 1, factor supplies satisfy the symmetry condition $\mathrm{D}_2 M(v_t, w_t) = \mathrm{D}_1 L(v_t, w_t)$. The elasticities (10)-(11) clearly imply that $g(m_t)K_t + f(n_t)$ is increasing in q_t, and then $\partial \mathrm{D}K_t/\partial q_t > 0$ follows immediately from (6). The result $\partial \mathrm{D}q_t/\partial K_t > 0$ is a consequence of (i) the fact that $\partial m_t/\partial K_t < 0$, and (ii) the fact that $\partial k_t/\partial K_t > 0$, because of $\partial v_t/\partial K_t > 0$ and the last part of Lemma 2.

To conclude the characterization of the static equilibrium conditions, it will be useful to note that $k_t = K_t/L(v_t, W(K_t, v_t))$ implies

$$\begin{bmatrix} \frac{K_t}{k_t}\frac{\partial k_t}{\partial K_t} & \frac{q_t}{k_t}\frac{\partial k_t}{\partial q_t} \end{bmatrix} = \begin{bmatrix} \frac{\frac{w_t}{\beta v_t + w_t} + \mathcal{E}_{L,w,t}}{\frac{w_t}{\beta v_t + w_t} + \left(1 + \mathcal{E}_{A,t}\right)\mathcal{E}_{L,w,t}} - \mathcal{E}_{L,v,t}\frac{K_t}{v_t}\frac{\partial v_t}{\partial K_t} & \frac{\left(\frac{\beta v_t}{\beta v_t + w_t}\right)\mathcal{E}_{L,w,t} - \left(\frac{w_t}{\beta v_t + w_t}\right)\mathcal{E}_{L,v,t}}{\frac{w_t}{\beta v_t + w_t} + \left(1 + \mathcal{E}_{A,t}\right)\mathcal{E}_{L,w,t}}\frac{q_t}{v_t}\frac{\partial v_t}{\partial q_t} \end{bmatrix}. \tag{12}$$

Assumption 2 says that $\mathcal{E}_{A,t} \geq 0$ and ensures that the elasticities of v_t with respect to K_t and q_t are both positive. It follows that the capital-labor ratio in the consumption sector is increasing in both K_t and q_t.

4.2 The Steady State

Steady states are defined by imposing $\mathrm{D}K_t = 0$ and $\mathrm{D}q_t = 0$ on the equilibrium conditions (6)-(8). The condition $\mathrm{D}K_t = 0$ says that $K = f(n)/(\delta - g(m))$, and $\mathrm{D}q_t = 0$ is the same as $qK = (1 - A(k))/(\rho + \delta - [g(m) - \mathrm{D}g(m)m])$. It will be convenient to write

$$n[m] = [\mathrm{D}f]^{-1}(\mathrm{D}g(m))$$

and define

$$m_\infty = \sup\{m : g(m) < \delta\}.$$

Since $g(m)$ is assumed to be strictly increasing, $m_\infty < \infty$ if and only if $g(m) \geq \delta$ for m large enough. In any case, the steady state managerial services needed for replicating capital, $mK = mf(n[m])/(\delta - g(m))$, explode as m approaches m_∞ from below.

What follows proves the existence and uniqueness of a steady state by first establishing the result for the Cobb-Douglas case, where $A(k)$ is a constant. The Cobb-Douglas economy implies a capital-labor ratio k that can be computed from (9), resulting in a map $\alpha \mapsto k$. The steady state for an economy with a non-constant $A(k)$ is simply a fixed point of the map $\alpha \mapsto A(k)$. Assumption 2 is enough to guarantee a unique fixed point.

4.2.1 The Cobb-Douglas Case

Suppose the labor share in the consumption sector is equal to $A(k) = \alpha \in (0, 1)$ identically. Construct the steady-state demand curve $D(\cdot)$ for managerial services by varying $m \in (0, m_\infty)$ and computing $(v, D(v))$ from

$$v = \frac{\delta - g(m)}{f(n[m])} \frac{(1-\alpha)\mathrm{D}g(m)}{\rho + \delta - [g(m) - \mathrm{D}g(m)m]}, \tag{13}$$

$$D(v) = \frac{mf(n[m])}{\delta - g(m)} + n[m]. \tag{14}$$

The residual supply curve $S(\cdot)$ is determined by

$$\alpha = (\beta v + w)L(v, w), \tag{15}$$

$$S(v) = M(v, w) - \beta L(v, w). \tag{16}$$

Taking into account the steady state conditions $\mathrm{D}K_t = 0$ and $\mathrm{D}q_t = 0$, (13) follows from $v = qK\mathrm{D}g(m)/K$ and (14) from $D(v) = mK + n$. Although $\mathrm{D}g(m)$ and $\rho + \delta - [g(m) - \mathrm{D}g(m)m]$ are both decreasing in m, an easy derivative calculation shows that the second factor in (13) is a strictly decreasing function of $m \in (0, m_\infty)$. It follows immediately that (13) defines v as a strictly decreasing function of $m \in (0, m_\infty)$. It is clear from (14) that $D(v)$ itself is strictly increasing in $m \in (0, m_\infty)$, and so (13)-(14) traces out a strictly decreasing demand curve for managerial services. The residual supply curve (15)-(16) is just a version of (4), with $A(k) = \alpha$ identically. As was shown in Lemma 3, it is strictly increasing.

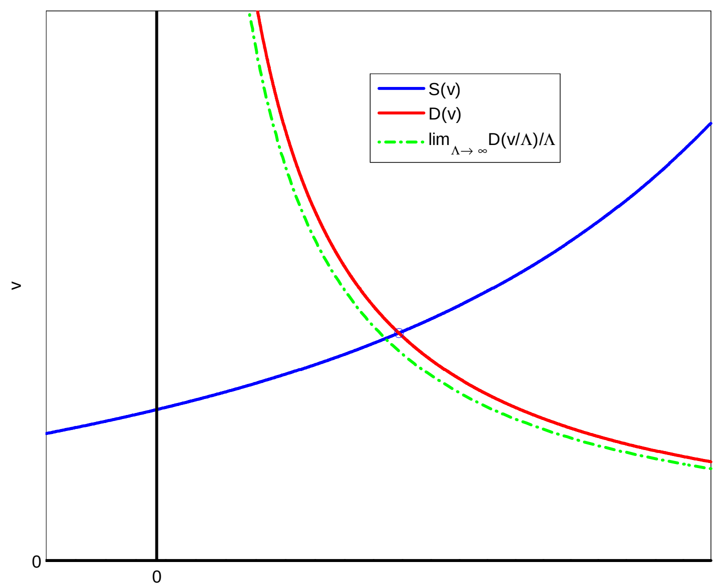

FIGURE 1 Steady State Demand and Supply for Managerial Services

Figure 1 shows an example. Since the supply and demand curves are strictly increasing and decreasing, respectively, it is immediate that the market clearing condition $D(v) = S(v)$ can have at most one solution, and so there can be at most one steady state. The existence of a steady state is guaranteed by the following proposition.

Proposition 1 *Given Assumptions 1 and 3 and a labor share $\alpha \in (0, 1)$, the Cobb-Douglas economy has a unique steady state, defined by (13)-(16), together with $K = f(n[m])/(\delta - g(m))$ and $qK = (1 - \alpha)/(\rho + \delta - (g(m) - \mathrm{D}g(m)m))$.*

Proof It suffices to show that the demand for managerial services is well defined for every $v \in (0, \infty)$. Consider what happens as $m \downarrow 0$. The assumption that g has a marginal product that ranges throughout $(0, \infty)$ implies that $\mathrm{D}g(m) \to \infty$. Efficiency requires that n declines as $m \downarrow 0$, and $g(m) - \mathrm{D}g(m)m \in (0, g(m))$ means that both $g(m)$ and $g(m) - \mathrm{D}g(m)m$ go to zero as $m \downarrow 0$. So the right-hand side of (13) goes to ∞ as $m \downarrow 0$. The demand (14) is certainly well defined when $m < m_\infty$, and so varying

$m \in (0, m_\infty)$ traces out a well defined demand curve for all v large enough. Next consider what happens as $m \uparrow m_\infty$. If $m_\infty < \infty$ then $\delta - g(m) \downarrow 0$ and $\mathrm{D}g(m) \downarrow \mathrm{D}g(m_\infty) \in (0, \infty)$. Since n increases with m, it also follows that $f(n)$ converges to a finite limit. So the right-hand side of (13) converges to 0. Alternatively, if $m_\infty = \infty$ then $\mathrm{D}g(m) \to 0$ as $m \to m_\infty$. And $\rho + \delta - [g(m) - \mathrm{D}g(m)m] > \rho$ ensures that the denominator of (13) is bounded away from zero. So the right-hand side of (13) converges to 0 again. That is, varying $m \in (0, m_\infty)$ near m_∞ traces out the demand curve for all v near 0. So varying $m \in (0, m_\infty)$ causes v to vary throughout $(0, \infty)$. This results in a well defined $D(v)$ for every such $v \in (0, \infty)$.$\blacksquare$

In preparation for the general case, consider how the steady state of a Cobb-Douglas economy depends on the labor share α. Note that an increase in α shrinks the demand curve $D(v)$ toward the quantity axis, essentially because it lowers the price of capital q, holding fixed K. At the same time, it is easy to see from (15) that an increase in α, holding fixed v, raises w. By (16), this reduces the household supply of managerial services and raises the supply of labor, lowering the residual supply $S(v)$ on both counts. So an increase in α moves the supply curve $S(v)$ toward the price axis. It follows that an increase in α lowers $D(v) = S(v)$. The definition (14) makes $D(v)$ an increasing function of m, and so a reduction in $D(v)$ has to go together with reduction in m. This lowers the steady state capital stock $K = f(n[m])/(\delta - g(m))$. Lemma A1 in the appendix shows that (15)-(16) entails that any increase in α that goes together with a decline in $S(v)$ implies an increase in $L(v, w)$. So an increase in α will not only lower the steady state capital stock but also raise the equilibrium supply of labor. This lowers the capital-labor ratio and thus proves the following lemma.

Lemma 5 *Given Assumptions 1 and 3, the steady state capital-labor ratio in the Cobb-Douglas economy is decreasing in the labor share α.*

4.2.2 The General Case

Now consider a more general production function F, with a labor share $A(k)$ that is non-decreasing in k. The conditions for a steady state are then the Cobb-Douglas conditions (13)-(16) together with the requirement that $\alpha = A(k)$, where $k = K/L(v, w)$ and $K = f(n[m])/(\delta - g(m))$. To set up a fixed point condition for the labor share, start with any

$\alpha \in (0,1)$ and use (13)-(16) to construct a unique steady state for the associated Cobb-Douglas economy. This can be done by Proposition 1. Take the associated (v, w, m) to compute

$$\alpha' = A(k), \quad k = \frac{f(n[m])/(\delta - g(m))}{L(v,w)}. \tag{17}$$

This produces a well defined mapping $\alpha \mapsto \alpha'$, from $(0,1)$ into $(0,1)$. If $\alpha' = \alpha$, then all the steady state equilibrium conditions for the general economy are satisfied. Proving the existence and uniqueness of a steady state now requires proving that $\alpha \mapsto \alpha'$ has precisely one fixed point. By Lemma 4, the mapping $\alpha \mapsto k$ implied by the Cobb-Douglas economy is decreasing. Assumption 2 requires $A(\cdot)$ to be weakly increasing. As a result, $\alpha \mapsto A(k) = \alpha'$ is a weakly decreasing map. Such a map can cross the line $\alpha' = \alpha$ only once. Because $\alpha \mapsto \alpha'$ is well defined for any $\alpha \in (0,1)$ and continuous, it follows that there is a unique fixed point.

Proposition 2 *Given Assumptions 1-3, the economy has a unique steady state.*

Constant elasticity of substitution production functions with an elasticity of substitution greater than 1 have $A(\cdot)$ decreasing instead of increasing. This was used by Jones and Manuelli [1990] to construct a one-sector economy without a steady state, with $k_t \to \infty$ and $A(k_t) \downarrow 0$ over time. Here, the mapping $\alpha \mapsto A(k) = \alpha'$ becomes increasing and it is no longer obvious that this mapping will have a fixed point. But F in this economy only determines $C_t = F(K_t, L(v_t, w_t))$ and output of new capital is the maximum of $g(m_t)K_t + f(n_t)$ subject to the constraint $m_t K_t + n_t \leq S(K_t, v_t)$. A production function F that generates long-run growth in Jones and Manuelli [1990] may very well result in a unique steady state here. A tractable example is discussed below, for a limiting economy in which $g(m) \uparrow \delta$.

4.2.3 The $g(m) \uparrow \delta$ Limit

If $\mathrm{D}f(0)$ is finite, contrary to Assumption 2, then the steady state may very well have $f(n) = 0$ together with $\delta = g(m)$ and an exponentially declining number of ever larger firms. Such a steady state would not be able to account for the fact that entry and exit rates in the US are around 11% and 10% per annum, or for the implied stability of the per-capita number of firms, or for the stability of the firm size distribution. But the tail index $\zeta \approx 1.10$ of the US size distribution of firms does suggest that

$\delta - g(m) = (g(m) - \delta_k)(\zeta - 1)$ must be very small. We therefore need to describe a scenario in which $\delta - g(m)$ is close to zero even though $f(n)$ is not.

Assume $m_\infty < \infty$, so that $g(m) > \delta$ is a technical possibility. Since m and n co-move, a small value of $\delta - g(m)$ inevitably means a large capital stock $K = f(n[m])/(\delta - g(m))$, and then $mK + n[m]$ implies large demand for managerial services. The supply of managerial services is bounded above by $M(\infty, 0) < \infty$ and so $g(m) \uparrow \delta$ is not possible unless the supply of managerial services can somehow be expanded.

To describe such a scenario, suppose the supplies of managerial and labor services are of the form

$$M(v, w) = \Lambda M_1(\Lambda v, \Lambda w), \quad L(v, w) = \Lambda L_1(\Lambda v, \Lambda w), \tag{18}$$

where M_1 and L_1 are baseline supply curves that satisfy Assumption 1, and Λ is a positive parameter. These factor supplies arise when the household characteristics (h_w, h_v) and h_v are replaced by $(\Lambda h_w, \Lambda h_v)$ for all households in the population. The household choice probabilities implied by $(\Lambda h_w, \Lambda h_v)$ are the same as they would be if (v, w) were replaced by $(\Lambda v, \Lambda w)$ without a change in household characteristics. But multiplying (h_w, h_v) by Λ also multiplies the supplies of managerial and labor services.

As before, it is useful to first consider the Cobb-Douglas case.

Proposition 3 *Suppose Assumptions 1 and 3 hold and F is Cobb-Douglas with labor share $\alpha \in (0, 1)$. Assume that $g(m) > \delta$ for m large enough and consider the factor supplies (18). Then $\Lambda[\delta - g(m)]$, Λv and Λw converge to limits in $(0, \infty)$ as Λ becomes large. As a consequence, $g(m) \uparrow \delta$ and $f(n) \uparrow f(n[m_\infty])$, implying that the capital stock diverges and that the tail index of the firm size distribution converges to 1.*

Proof Recall the Cobb-Douglas steady state conditions (13)-(16) and let S_1 be residual supply curve of managerial services implied by M_1 and L_1 and (15)-(16). Write $u = \Lambda v$ and observe that the steady state equilibrium condition $S(u/\Lambda) = D(u/\Lambda)$ can be written as $S_1(u) = D(u/\Lambda)/\Lambda$. Fix some $u \in (0, \infty)$ and replace the left-hand side of (13) by $v = u/\Lambda$. Letting $\Lambda \to \infty$ then forces $m \uparrow m_\infty$, no matter what the value of $u \in (0, \infty)$. In other words, (13) implies

$$\lim_{\Lambda \to \infty} \Lambda[\delta - g(m)] = \left(\frac{(1 - \alpha)\mathrm{D}g(m_\infty)}{\rho + \mathrm{D}g(m_\infty)m_\infty]} \right)^{-1} f(n[m_\infty])u. \tag{19}$$

The right-hand side of this equation is well defined and in $(0, \infty)$. From (14), the limiting scaled demand for managerial services is then

$$D_\infty(u) = \lim_{\Lambda \to \infty} D(u/\Lambda)/\Lambda = \frac{\mathrm{D}g(m_\infty)m_\infty}{\rho + \mathrm{D}g(m_\infty)m_\infty} \frac{1-\alpha}{u}. \tag{20}$$

This hyperbola is guaranteed to intersect the strictly increasing supply curve $S_1(u)$ precisely once, and so the limiting market clearing condition $S_1(u) = \lim_{\Lambda \to \infty} D(u/\Lambda)/\Lambda$ has a well defined and unique solution for the scaled factor price u.■

As the proof of Proposition 3 shows, the scaled demand curve $D(u/\Lambda)/\Lambda$ for managerial services is unit elastic in the large-Λ limit. In other words, $D(v)$ is close to unit elastic for small v.[3] The underlying reason is that $qK = (1-\alpha)/(\rho + \delta - [g(m) - \mathrm{D}g(m)m])$ and thus $vK = qK\mathrm{D}g(m)$ converge to well defined limits as $m \uparrow m_\infty$, while $K = f(n)/(\delta - g(m))$ diverges. This means that $D(v) = mK + n$ is dominated by mK and that $K = qK\mathrm{D}g(m)/v$ behaves like $1/v$ when m is close to m_∞. It turns out that the demand curve $D(v)$ can be quite close to a hyperbola even when $g(m)/\delta$ is well below 1. An illustration is provided by the example given in Figure 1, which has $\delta = 0.10$, $g(m) \approx \delta/2$, and $m \approx 0.22 m_\infty$.

Notice that the proof of Proposition 3 does not depend directly on the labor supply curve $L(v, w)$, only via the residual supply of managerial services. One can keep $L(v, w) = L_1(v, w)$ and instead replace β by $\Lambda\beta$ to obtain a residual supply of managerial services of the form used in the proof of Proposition 3. The interpretation is that managerial services improve with $\Lambda > 1$ only to the extent that they are used to replicate, and not when used to form production teams with labor. In that case, Λv and w converge to limits in $(0, \infty)$, and the capital-labor ratio used to produce consumption diverges. Since the Cobb-Douglas technology has a constant labor share, this exploding capital-labor ratio does not alter the incentives to produce and replicate capital.

With $A(k)$ increasing, an exploding capital-labor ratio k would drive the factor share $1 - A(k)$ of capital down to zero unless $A(\infty) < 1$ (that is to say, unless F is approximately Cobb-Douglas for large capital-labor ratios.) This would upset the argument

[3] More precisely, $\lim_{v \downarrow 0} D(v)v \in (0, \infty)$. This implies but is not implied by $\lim_{v \downarrow 0} D(\lambda v)/D(v) = 1/\lambda$, which says that $D(v)$ is regularly varying of degree -1 near zero (Bingham, Goldie and Teugels [1987, p.21]).

used in Proposition 3 to prove that the derived demand for managerial services is an approximate hyperbola. Maintain, therefore, the factor supplies described in (18).

Proposition 4 *Suppose Assumptions 1-3 hold and that $g(m) > \delta$ for m large enough, and that the factor supplies are of the form (18). Then the conclusions of Proposition 3 apply.*

Proof This uses the fact that the limiting economy obtained by letting $\Lambda \to \infty$ will also have a unique steady state that is a fixed point of a mapping constructed as in Proposition 2 based on the limiting Cobb-Douglas economy described in Proposition 3. The details are somewhat intricate and given in the appendix.■

Consider the Cobb-Douglas economy with some $\alpha \in (0,1)$. The limit described in Proposition 3 says that the capital-labor ratio used to produce consumption will satisfy

$$\lim_{\Lambda\to\infty} k = \lim_{\Lambda\to\infty} \frac{f(n[m])/L_1(\Lambda v, \Lambda w)}{\Lambda(\delta - g(m))} \in (0,\infty),$$

since $m \to m_\infty$ and $\Lambda(\delta - g(m))$, Λv and Λw converge to well-defined limits in $(0,\infty)$. The factor supplies (18) are such that both $L(v,w)$ and $M(v,w)$ become large as Λ becomes large. This stabilizes the capital-labor ratio used to produce consumption. In turn this stabilizes the mapping $\alpha \mapsto A(k) = \alpha'$, and then Proposition 4 follows.

Robust Entry Note that n increases to the positive limit $n[m_\infty]$ as the supply of managerial services becomes large. So there will be robust entry in the limit, and the number of firms converges to $f(n[m_\infty])/\delta$. But because the average size of firms diverges, the contribution of new firms to the aggregate accumulation of capital will be negligible—both $n/(mK+n)$ and $f(n)/(g(m)K+f(n))$ converge to zero as Λ becomes large. A large-Λ economy is an economy in which the managerial resources needed to produce capital are abundant. Entry is one way in which these resources can be used, but the marginal product of $f(n)$ converges to 0 as n becomes large. On the other hand, the marginal product of $g(m)$ is bounded away from 0 on the interval $[0, m_\infty)$. In a steady state, this directs sufficiently abundant managerial resources mostly into the process of replicating existing capital.

4.2.4 A Counterexample

By Proposition 2, the assumption that the labor share $A(k)$ is increasing in k is sufficient to guarantee existence of a steady state. But a decreasing $A(k)$ is not necessary for existence of a unique steady state. The limiting economy can be used to construct a transparent counterexample.

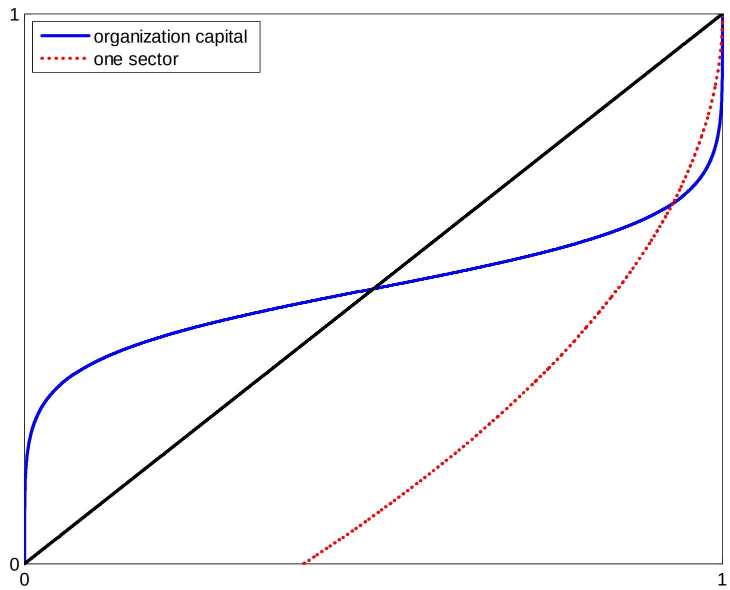

FIGURE 2 Existence When $A(k)$ is Decreasing

Simplify by taking $\beta = 0$ and assuming that the large-Λ factor supplies implied by (18) are determined by supply curves $L_1(v, w) = \mathcal{L}(w)$ and $M_1(v, w) = \mathcal{M}(v)$.[4] That is, managers are not needed to produce consumption, and there are separate populations of households who supply only labor or only managerial services. In the Cobb-Douglas version of such an economy, steady state wages are determined by $\alpha = w\mathcal{L}(w)$ and the supply of managerial services available to produce capital is simply $\mathcal{M}(v)$. The limiting demand curve $D_\infty(v)$ for managerial services is defined in (19)-(20). It is the hyperbola $vD_\infty(v) = (1 - \alpha)X_\infty$, where $X_\infty = \mathrm{D}g(m_\infty)m_\infty/(\rho + \mathrm{D}g(m_\infty)m_\infty)$. So the market clearing condition for managerial services is simply $(1 - \alpha)X_\infty = v\mathcal{M}(v)$, which makes v a decreasing function of α. The resulting capital stock is then $K = \mathcal{M}(v)/m_\infty$. In this setting, it is easy to verify that the elasticity of the capital-labor ratio $k = K/\mathcal{L}(w)$

[4] Note well that the arguments of these supply curves are the large-Λ limiting values of the scaled factor prices $(v, w) = (\widetilde{v}, \widetilde{w})\Lambda/C$.

with respect to α satisfies

$$(1-\alpha) \times \frac{\alpha}{k}\frac{\partial k}{\partial \alpha} = -\left(\alpha \times \frac{\mathcal{E}_{M,v}}{1+\mathcal{E}_{M,v}} + (1-\alpha) \times \frac{\mathcal{E}_{L,w}}{1+\mathcal{E}_{L,w}}\right).$$

The key observation is that the right-hand side must be in $(-1, 0)$. In turn, the concavity of F requires that $-\mathrm{D}A(k)k/A(k) \leq 1 - A(k)$.[5] At a steady state, $\alpha = A(k)$, and so the elasticity of $\alpha \mapsto A(k) = \alpha'$ is smaller than 1 in any steady state. This rules out the possibility of multiple steady states, even when $A(k)$ is a decreasing function.

Figure 2 shows an example in which there is a steady state, even though $A(k)$ decreases monotonically and converges to 0 as k becomes large. As argued, this steady state is unique in $(0, 1)$. The production function F is a CES production function with an elasticity equal to 2, and labor and managerial services supplies have unit elasticities.

For comparison, Figure 2 also includes the mapping $\alpha \mapsto \alpha'$ for a Cass-Koopmans economy with the same CES production function. In such an economy, the steady state capital-labor ratio k must solve $\rho + \delta = \mathrm{D}_1F(k, 1)$. The mapping $\alpha \mapsto \alpha'$ can then be defined by $\alpha' = A(k[\alpha])$ with $k[\alpha]$ solving $(\rho + \delta)/F(1, 1/k[\alpha]) = \alpha$. An elasticity of substitution greater than 1 implies $F(1, 0) > 0$, and so $k[\alpha]$ is not well defined for any α close enough to 0, which is what rules out a fixed point and generates growth in Jones and Manuelli [1990].

4.2.5 A Note on Related Work

The interpretation of Zipf's law implied by Propositions 3 and 4 is closely related to the one given in Luttmer [2011]. But there the human factors of production that are used to produce consumption and replicate capital are perfect substitutes, while here the human factors of production that are used to replicate capital and create capital from scratch are perfect substitutes. It is clear that one can generalize and take all these human inputs to be imperfect substitutes in production.

5. The Phase Diagram

The properties of the equilibrium trajectory of (K_t, q_t) starting from some initial $K_0 \geq 0$ can be characterized using a simple phase diagram. For this we need to first characterize

[5] The elasticity of $A(k)$ is equal to the curvature of $F(1, l)$ evaluated at $l = 1/k$, minus $1 - A(k)$.

the (K, q) curves associated with $\mathrm{D}K_t = 0$ and $\mathrm{D}q_t = 0$.

5.1 The Conditions $\mathbf{D}K_t = 0$ and $\mathbf{D}q_t = 0$

The condition $\mathrm{D}K_t = 0$ corresponds to

$$K = \frac{f(n[m])}{\delta - g(m)} \tag{21}$$

together with the market clearing condition $mK + n[m] = S(K, q\mathrm{D}g(m))$ for managerial services. Since $\mathrm{D}_1 S(K, v) \leq 0$ and $\mathrm{D}_2 S(K, v) > 0$, this defines an increasing function $K \mapsto q$. The condition $\mathrm{D}q_t = 0$ is determined by

$$q = \frac{(1 - A(k))/K}{\rho + \delta - [g(m) - \mathrm{D}g(m)m]} \tag{22}$$

together with the market clearing condition $mK + n[m] = S(K, q\mathrm{D}g(m))$ for managerial services, as well as $k = K/L(v, W(K, v))$ and $v = q\mathrm{D}g(m)$. This implies a well defined function $q \mapsto K$. To see this, hold fixed some $q \in (0, \infty)$ and note that the market clearing condition for managerial services produces m as a decreasing function of K. As a result, $1/(\rho + \delta - [g(m) - \mathrm{D}g(m)m])$ is a decreasing function of K as well. The consumption sector revenues $(1 - A(k))/K$ per unit of capital are also decreasing in K, because of Lemma 2 and the fact that $v = q\mathrm{D}g(m)$ makes v and increasing function of K. And these revenues vary throughout $(0, \infty)$ with K. It follows that there is a unique K that satisfies (22) for any given q.

5.2 The Equilibrium Trajectory

From Proposition 1, we know that the curve $K \mapsto q$ defined by $\mathrm{D}K_t = 0$ and the curve $q \mapsto K$ defined by $\mathrm{D}q_t = 0$ intersect precisely once. Figure 2 shows the phase diagram, with K on the horizontal axis and q on the vertical axis. Lemma 4 implies that $\mathrm{D}K_t > 0$ above the $\mathrm{D}K_t = 0$ curve and $\mathrm{D}K_t < 0$ below that curve. Lemma 4 also implies that $\mathrm{D}q_t > 0$ to the right of the $\mathrm{D}q_t = 0$ curve and $\mathrm{D}q_t < 0$ to its left. Included in Figure 2 is the downward sloping stable manifold—the only trajectory for (K_t, q_t) that converges to the steady state. Trajectories for which q_t reaches zero can be ruled out because $\mathrm{D}q_t < 0$ at that point, so that q_t could then not remain non-negative. Combining (6)-(7) shows that the discounted value of the aggregate capital stock evolves according to

$$\mathrm{D}[e^{-\rho t} q_t K_t] = e^{-\rho t} q_t f(n_t) + e^{-\rho t} q_t K_t \mathrm{D}g(m_t) m_t - e^{-\rho t}(1 - A(k_t)). \tag{23}$$

Trajectories for which q_t diverges and K_t converges therefore result in $e^{-\rho t} q_t K_t$ growing at the approximate rate $\mathrm{D}g(m_t)m_t \geq 0$ in the long run. This violates the transversality condition for this economy. With the transversality condition imposed, $q_t K_t$ will be the present value of the aggregate product market profits minus the cost of keeping the capital stock on its equilibrium path: the total cost of producing capital from scratch, and the cost of managerial services used in replication.

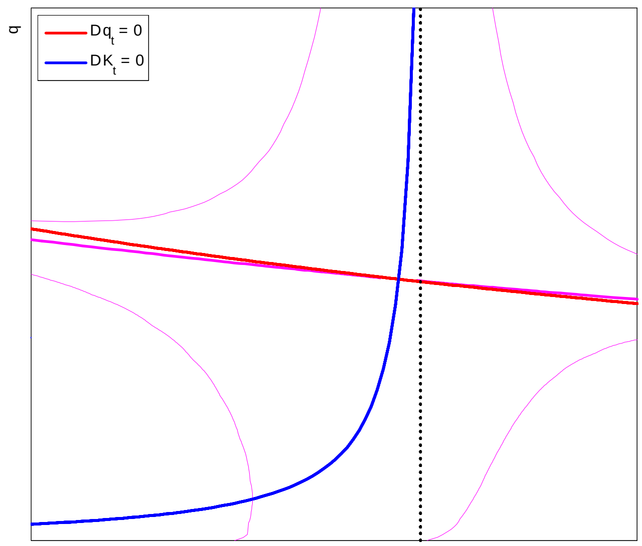

FIGURE 3 The Phase Diagram

The static equilibrium conditions (8) imply $m_t = \mathcal{M}(K_t, q_t)$ and $n_t = \mathcal{N}(K_t, q_t)$, and these functions are both decreasing in K_t and increasing in q_t. Since the stable manifold is downward sloping, it follows that, along the equilibrium trajectory, the equilibrium values of m_t and n_t are decreasing functions of the capital stock. Both inputs will be above their steady state levels when the capital stock is below its steady state. As a result, the growth rate $\mathrm{D}K_t/K_t = -\delta + g(m_t) + f(n_t)/K_t$ will be high when K_t is low, and decline monotonically towards zero over time. But what happens to gross investment $g(m_t)K_t + f(n_t)$ (measured in units of capital, not consumption) when K_t is below its steady state is not clear: the direct effect of a low K_t may dominate potentially small increases in $g(m_t)$ and $f(n_t)$.

5.3 To Create or Replicate

More can be said about the share $n_t/(m_tK_t + n_t) = n_t/S(q_t\mathrm{D}f(n_t))$ of managerial services used for entry and the related share of entry in aggregate gross investment $f(n_t)/(g(m_t)K_t + f(n_t))$. For example, if f and g are both Cobb-Douglas production functions with the same factor share parameter, then the efficiency condition $1 = \mathrm{D}f(n_t)/\mathrm{D}g(m_t)$ implies that the ratios n_t/m_t and $f(n_t)/g(m_t)$ are constant. This then immediately implies that the contribution of entrants declines when the capital stock increases. More generally, holding fixed q_t, the ratio $n_t/S(q_t\mathrm{D}f(n_t))$ is increasing in n_t. The static equilibrium conditions imply that n_t itself is decreasing in K_t, holding fixed q_t, and hence so is the ratio $n_t/S(q_t\mathrm{D}f(n_t))$. The effect of q_t on this ratio is ambiguous: the static equilibrium conditions imply that both n_t and $v_t = q_t\mathrm{D}f(n_t)$ are increasing in q_t. A precise statement is given in the following proposition.

Proposition 5 *The stable manifold is downward sloping. For all (K_t, q_t) on the stable manifold,*

$$\begin{aligned}
\frac{1}{\mathcal{C}_{f,t}} &> \frac{1}{\mathcal{C}_{g,t}} \Rightarrow \frac{\mathrm{d}}{\mathrm{d}K_t}\frac{n_t}{m_tK_t + n_t} < 0,\\
\frac{\mathcal{S}_{f,t}}{\mathcal{C}_{f,t}} &> \frac{\mathcal{S}_{g,t}}{\mathcal{C}_{g,t}} \Rightarrow \frac{\mathrm{d}}{\mathrm{d}K_t}\frac{f(n_t)}{g(m_t)K_t + f(n_t)} < 0,
\end{aligned}$$

where m_t and n_t are determined by the static equilibrium conditions (8).

In the Cobb-Douglas case, curvature equals 1 minus the factor share, and then Proposition 3 says that the contribution of entry declines as the economy approaches the steady state from below if the managerial factor share of f is greater than that of g. Far below the steady state, "everyone" is an entrepreneur setting up new businesses, and over time the growth of successful businesses becomes more important.

5.4 The Speed of Convergence

Near the steady state, the speed at which the economy converges is determined by the matrix of derivatives $\partial[\mathrm{D}K_t, \mathrm{D}q_t]/\partial[K_t, q_t]$ implied by (6)-(8), evaluated at the steady state. In computing these derivatives, it is useful to note that (7) yields

$$\begin{aligned}\left(\frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t}\right)_{(q_t,k_t,m_t)\text{ fixed}} &= \rho+\delta-(1-\mathcal{S}_g)g(m),\\ \left(\frac{\partial \mathrm{D}q_t}{\partial q_t}\right)_{(K_t,k_t,m_t)\text{ fixed}} &= \rho+\delta-(1-\mathcal{S}_g)g(m),\\ \left(\frac{k_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial k_t}\right)_{(K_t,q_t,m_t)\text{ fixed}} &= (\rho+\delta-(1-\mathcal{S}_g)g(m))\times\frac{A(k)\mathcal{E}_A}{1-A(k)}.\end{aligned} \tag{24}$$

In particular, the first of these elasticities equals $(1-A(k_t))/(q_tK_t)$, and then the steady state condition $\mathrm{D}q_t=0$ gives the above result. The partial elasticities (24) highlight the important role of the managerial factor share $\mathcal{S}_g$, especially in the $g(m)\uparrow\delta$ limit. Completing the computation of $\partial[\mathrm{D}K_t,\mathrm{D}q_t]/\partial[K_t,q_t]$ gives

$$\begin{aligned}\begin{bmatrix}\frac{\partial \mathrm{D}K_t}{\partial K_t} & \frac{q_t}{K_t}\frac{\partial \mathrm{D}K_t}{\partial q_t}\\ \frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t} & \frac{\partial \mathrm{D}q_t}{\partial q_t}\end{bmatrix} &= \begin{bmatrix}-(\delta-g(m)) & 0\\ \rho+\delta-g(m) & \rho+\delta-g(m)\end{bmatrix}\\ &+\mathcal{S}_g g(m)\begin{bmatrix}\frac{K}{m}\frac{\partial m}{\partial K}+\frac{n}{mK}\frac{K}{n}\frac{\partial n}{\partial K} & \frac{q}{m}\frac{\partial m}{\partial q}+\frac{n}{mK}\frac{q}{n}\frac{\partial n}{\partial q}\\ 1-\mathcal{C}_g\times\frac{K}{m}\frac{\partial m}{\partial K} & 1-\mathcal{C}_g\times\frac{q}{m}\frac{\partial m}{\partial q}\end{bmatrix}\\ &+(\rho+\delta-(1-\mathcal{S}_g)g(m))\left(\frac{A(k)\mathcal{E}_A}{1-A(k)}\right)\begin{bmatrix}0 & 0\\ \frac{K}{k}\frac{\partial k}{\partial K} & \frac{K}{k}\frac{\partial k}{\partial q}\end{bmatrix}\end{aligned} \tag{25}$$

The direct effects (24) are easy to recognize, and the important role of $\delta-g(m)$ and $\mathcal{S}_g$ is apparent. The partial elasticities on the right-hand side of (25) are the elasticities implied by the static equilibrium conditions given in (10), (11) and (12). The third line in (25) vanishes in the Cobb-Douglas case, where $\mathcal{E}_A=0$. Recall from Lemma 4 that the off-diagonal elements of (25) are positive. At the steady state, $\delta-g(m)$ must be positive, and then (10)-(11) implies that $\partial\mathrm{D}K_t/\partial K_t<0$. Lemma 4 then implies that $\partial\mathrm{D}q_t/\partial q_t>0$. With only $\partial\mathrm{D}K_t/\partial K_t$ negative, it follows that the determinant of the matrix (25) is negative, and so its eigenvalues must be of opposite signs. Near the steady state, a linear approximation for the stable manifold is obtained by selecting q_t so that $[K_t-K,q_t-q]$ is an eigenvector associated with the negative eigenvalue. The absolute value of the negative eigenvalue then measures the exponential rate at which the state converges to the steady state. A calculation shows that this speed is given by

$$\text{speed}=-\frac{\rho}{2}+\sqrt{\left(\frac{\rho}{2}\right)^2+\mathcal{D}} \tag{26}$$

where $\mathcal{D}$ is the interesting part of the discriminant,

$$\mathcal{D} = \frac{\partial \mathrm{D}K_t}{\partial q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t} - \frac{\partial \mathrm{D}K_t}{\partial K_t}\left(\rho - \frac{\partial \mathrm{D}K_t}{\partial K_t}\right). \tag{27}$$

The associated eigenvector implies that the elasticity of q_t with respect to K_t along the stable manifold is equal to

$$\frac{K_t}{q_t}\frac{\partial q_t}{\partial K_t} = -\frac{\frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t}}{\frac{\rho}{2} - \frac{\partial \mathrm{D}K_t}{\partial K_t} + \sqrt{\left(\frac{\rho}{2}\right)^2 + \mathcal{D}}}, \tag{28}$$

at the steady state. As expected from Figure 2, this slope is unambiguously negative.

It is clear from (26)-(27) that the economy converges slowly when $\mathcal{D}$ is small. To a first-order approximation, the speed (26) is then simply $\mathcal{D}$. The approximate speed $\mathcal{D}$ is increasing in $-\partial \mathrm{D}K_t/\partial K_t$. This derivative is just the measure of mean reversion for K_t that applies when the price q_t is held fixed. The extent to which changes in the price of capital can speed up convergence is measured by the off-diagonal slopes $\partial \mathrm{D}K_t/\partial q_t > 0$ and $\partial \mathrm{D}q_t/\partial K_t > 0$. When K_t is low, it helps if $g(m_t)K_t + f(n_t)$ responds elastically to a high q_t. In turn, as indicated by the slope (28) of the stable manifold, a high q_t is more likely when $\partial \mathrm{D}q_t/\partial K_t > 0$ is large. Holding fixed future prices, the current price of capital will then be particularly high when K_t is low. The size of this effect is large if the profits $g(m_t) - \mathrm{D}g(m_t)m_t + (1 - A(k_t))/K_t$ per unit of capital are particularly high when K_t is low.

5.4.1 The Basic Mechanism Again

The speed calculation given in the Introduction and Section 3 survives if the factor share parameter $\mathcal{S}_g$ is close to zero and $A(k)$ is constant—the Cobb-Douglas case. The derivative matrix (25) then simplifies drastically, to

$$\begin{bmatrix} \frac{\partial \mathrm{D}K_t}{\partial K_t} & \frac{\partial \mathrm{D}K_t}{\partial q_t} \\ \frac{\partial \mathrm{D}q_t}{\partial K_t} & \rho - \frac{\partial \mathrm{D}K_t}{\partial K_t} \end{bmatrix} \approx \begin{bmatrix} -(\delta - g(m)) & 0 \\ \rho + \delta - g(m) & \rho + \delta - g(m) \end{bmatrix}.$$

Clearly, this matrix has eigenvalues $-(\delta - g(m)) < 0$ and $\rho + \delta - g(m) > 0$, and so the speed of convergence is simply $\delta - g(m)$. The evidence on the US firm size distribution suggests that this is a very small number.

5.4.2 Inelastic Managerial Services

It is instructive to directly calculate the speed of convergence when, contrary to Assumption 1, the residual supply of managerial services is completely inelastic at some positive $S(\infty)$. These inelastically supplied managerial services are only good for replicating capital or producing new capital from scratch. The dynamics of the capital stock must therefore be

$$\mathrm{D}K_t = -\delta K_t + \max_{m,n} \{g(m)K_t + f(n) : mK_t + n \leq S(\infty)\}. \tag{29}$$

The unique steady state capital stock is the maximal sustainable steady state capital stock of this economy. Because the residual supply of managerial services cannot respond, the price q_t plays no role in (29) and the only choice is how much of the fixed supply of managerial services to allocate to replication and entry. The constraint on managerial services immediately forces

$$\frac{K_t}{m_t}\frac{\partial m_t}{\partial K_t} + \frac{n_t}{m_t K_t}\frac{K_t}{n_t}\frac{\partial n_t}{\partial K_t} = -1. \tag{30}$$

Setting $\mathcal{E}_{S,v} = 0$ in (10) gives the same result, of course. Totally differentiating $g(m_t)K_t + f(n_t)$ with respect to K_t, imposing the efficiency condition $\mathrm{D}g(m_t) = \mathrm{D}f(n_t)$, and using (30) yields

$$\begin{aligned} \frac{\partial \mathrm{D}K_t}{\partial K_t} &= -\delta + g(m_t) - \mathrm{D}g(m_t)m_t \\ &= -\left[\delta - (1 - \mathcal{S}_{g,t})\, g(m_t)\right] < -\left[\delta - g(m_t)\right]. \end{aligned} \tag{31}$$

The fact that m_t and n_t do not respond to q_t implies that $\partial \mathrm{D}K_t/\partial q_t = 0$, and then (31) also follows readily from the general formula implied by (25) and (26). The term $g(m_t) - \mathrm{D}g(m_t)m_t = (1 - \mathcal{S}_{g,t})\, g(m_t)$ that appears in (31) is simply the marginal product of capital of the production function $G(K, M) = Kg(M/K)$ evaluated at $(K, M) = (K_t, m_t K_t)$. If $\mathrm{D}K_t = -\delta K_t + G(K_t, S(\infty) - n) + f(n)$ for some fixed $n \in (0, S(\infty))$, then (31) is immediate. Optimally allocating managerial services across the replication and entry activities has a second order effect on the speed of convergence.

The convergence rate (31) highlights the role of the managerial factor share $\mathcal{S}_{g,t}$. A higher managerial factor share $\mathcal{S}_{g,t}$ raises the speed of convergence, subject to a speed limit given by δ. When $\delta - g(m_t) > 0$ is small, the approximate convergence rate is simply $\delta \mathcal{S}_{g,t}$. The economy will converge much more slowly than the speed limit δ if the factor share of managerial services in replication is small.

5.4.3 Abundant Managerial Services

But assuming a completely inelastic residual supply of managerial services is extreme, given $\beta > 0$, and given that household members cannot only withdraw from the labor market but also switch between supplying labor and supplying managerial services. Return, therefore, to the economy in which Assumption 1 holds. It is not difficult to verify using (25) that the expression (27) for the approximate speed of convergence is an increasing function of $\delta - g(m) > 0$, viewed as an independent parameter. The following describes what can be said in the $g(m) \uparrow \delta$ limit computed in Propositions 3 and 4.

In the $g(m) \uparrow \delta$ limit, the terms $\delta - g(m)$ disappear from (25), and mK becomes large relative to n. So the coefficient $n/(mK)$ in (25) that multiplies the elasticities of n with respect to K and q goes to zero. The matrix of elasticities of (m, n) with respect to (K, q) implied by (10)-(11) converges to

$$\begin{bmatrix} \frac{K}{m}\frac{\partial m}{\partial K} & \frac{q}{m}\frac{\partial m}{\partial q} \\ \frac{K}{n}\frac{\partial n}{\partial K} & \frac{q}{n}\frac{\partial n}{\partial q} \end{bmatrix} = \begin{bmatrix} 1/\mathcal{C}_g \\ 1/\mathcal{C}_f \end{bmatrix} \frac{\begin{bmatrix} \mathcal{E}_{S,K} - 1 & \mathcal{E}_{S,v} \end{bmatrix}}{\mathcal{E}_{S,v} + 1/\mathcal{C}_g}. \tag{32}$$

As long as $\mathcal{C}_f \in (0, \infty)$, the managerial services allocated to entry respond to changes in the state (K, q), even in the $g(m) \uparrow \delta$ limit. Since the stable manifold is downward sloping, this manifests itself in more entry when K is below its steady state. But the entry elasticities disappear from (25) because the contribution of new capital created from scratch becomes negligible. The only elasticities that matter for the speed of convergence are those of m with respect to (K, q). As long as $\mathcal{E}_{S,v} > 0$, both these elasticities will be small when the curvature of $g(\cdot)$ is high. Holding fixed q, a low K may then raises m a bit, but the resulting steep reduction of the marginal product $\mathrm{D}g(m)$ implied by a high $\mathcal{C}_g$ mostly reduces the supply of managerial services.

Using (32) together with $g(m) \uparrow \delta$ to simplify (25) yields

$$\begin{aligned} \begin{bmatrix} \frac{\partial \mathrm{D}K_t}{\partial K_t} & \frac{q_t}{K_t}\frac{\partial \mathrm{D}K_t}{\partial q_t} \\ \frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t} & \frac{\partial \mathrm{D}q_t}{\partial q_t} \end{bmatrix} \quad = \quad & \begin{bmatrix} 0 & 0 \\ \rho & \rho \end{bmatrix} + \frac{\mathcal{S}_g \delta}{1 + \mathcal{C}_g \mathcal{E}_{S,v}} \begin{bmatrix} \mathcal{E}_{S,K} - 1 & \mathcal{E}_{S,v} \\ 1 + \mathcal{C}_g(1 + \mathcal{E}_{S,v} - \mathcal{E}_{S,K}) & 1 \end{bmatrix} \\ & + (\rho + \mathcal{S}_g \delta) \left(\frac{A(k)\mathcal{E}_A}{1 - A(k)} \right) \begin{bmatrix} 0 & 0 \\ \frac{K}{k}\frac{\partial k}{\partial K} & \frac{K}{k}\frac{\partial k}{\partial q} \end{bmatrix}. \end{aligned} \tag{33}$$

Again, the last line is zero in the Cobb-Douglas case. A bit of algebra shows that the implied approximate speed of convergence (27) is now

$$\mathcal{D} = \frac{(\rho + \mathcal{S}_g \delta)\, \mathcal{S}_g \delta}{\mathcal{C}_g + \frac{1-\mathcal{C}_g}{\mathcal{E}_{S,v}+1}} \times \mathcal{F} \tag{34}$$

where

$$\mathcal{F} = \frac{1 - \mathcal{E}_{S,K} + \left(1 + \frac{A(k)\mathcal{E}_A}{1-A(k)} \frac{K}{k} \frac{\partial k}{\partial K}\right) \mathcal{E}_{S,v}}{1 + \mathcal{E}_{S,v}} + \frac{\delta \mathcal{S}_g}{\rho + \delta \mathcal{S}_g} \frac{(1 - \mathcal{E}_{S,K})(-\mathcal{E}_{S,K})}{(1 + \mathcal{C}_g \mathcal{E}_{S,v})(1 + \mathcal{E}_{S,v})}. \tag{35}$$

The residual managerial supply elasticities $\mathcal{E}_{S,K}$ and $\mathcal{E}_{S,v}$ were defined in Lemma 3 and $(K/k)\partial k/\partial K$, the elasticity of k with respect to K implied by the static equilibrium conditions (holding fixed q), was reported in (12). Recall that Assumption 2 ensures that $\mathcal{E}_{S,K} \leq 0$ and $\mathcal{E}_A \geq 0$, and both elasticities are zero if the production function in the consumption sector is Cobb-Douglas. Also, (12) implies that $(K/k)\partial k/\partial K$ is positive. So the assumption that the labor share $A(k)$ is non-decreasing in the consumption-sector capital-labor ratio implies $\mathcal{F} \geq 1$.

The Cobb-Douglas case yields $\mathcal{F} = 1$, and so the first factor on the right-hand side of (34) is the approximate speed for a Cobb-Douglas economy. The residual managerial services price elasticity $\mathcal{E}_{S,v}$ depends not only on $\mathcal{E}_A$ but also on all the factor supply elasticities. So one can consider varying $\mathcal{E}_A$ while holding fixed $\mathcal{E}_{S,v}$. For such comparisons, economies with a strictly increasing labor share function $A(k)$ converge more quickly than the corresponding Cobb-Douglas economy. The simple intuition is that $\mathcal{E}_A > 0$ and $\partial k/\partial K > 0$ has the effect of raising profits per unit of capital when K is low, thus increasing the incentives to accumulate organization capital.

More on the Cobb-Douglas Case The leading factor in (34) formula immediately confirms that a low value of $\mathcal{S}_g$ leads to slow convergence. Note also that $\mathcal{C}_g = 1$ implies an exact speed equal to $\delta\mathcal{S}_g$, independently of $\mathcal{E}_{S,v}$. The economy converges more quickly than $\delta\mathcal{S}_g$ when $\mathcal{C}_g \in (0, 1)$, as is the case if $g(m)$ is also Cobb-Douglas, so that $\mathcal{C}_g = 1 - \mathcal{S}_g$. The right-hand side of (34) is then an increasing function of $\mathcal{E}_{S,v}$. The speed of convergence declines towards $\delta\mathcal{S}_g$, as in (31), when $\mathcal{E}_{S,v} \downarrow 0$. As before, slow convergence requires a small managerial services share $\mathcal{S}_g$.

A different slow convergence scenario arises when $\mathcal{C}_g \to \infty$. The maintained assumption that $\mathcal{E}_{S,v} > 0$ ensures that the supply of managerial services can respond to a low factor v_t. More generally, the speed of convergence is a decreasing function of $\mathcal{E}_{S,v} > 0$

whenever $\mathcal{C}_g > 1$. In that case, as $\mathcal{E}_{S,v}$ becomes large, the speed of convergence converges to a positive limit that is strictly less than $\delta\mathcal{S}_g$, and much less than $\delta\mathcal{S}_g$ if $\mathcal{C}_g$ is large. The combination of $\mathcal{C}_g > 1$ and a large managerial supply elasticity $\mathcal{E}_{S,v}$ leads to slow convergence rates. This is intuitive. When K_t is low, a recovery could be helped along by a substantial increase in m_t. But $\mathcal{C}_g > 1$ implies that this would have a strong negative effect on the marginal product $\mathrm{D}g(m_t)$, and, when $\mathcal{E}_{S,v,t}$ is large, a particularly strong negative effect on the supply of managerial services. The fact that managers have other options limits the extent to which m_t can rise when K_t falls below the steady state.

It bears repeating that this is not an economy without entry. But when $g(m) \uparrow \delta$, how entry responds to the state of the economy has a negligible effect on aggregate dynamics.

5.5 Relation to Cass-Koopmans

It is interesting to see how these results relate to a Cass-Koopmans economy with logarithmic utility and a labor supply implied by assuming that labor and managerial services are perfect substitutes in production. In such an economy, $F(K, L) = Kg(L/K)$ and the capital stock evolves according to $\mathrm{D}K_t = (g(m_t) - \delta)K_t - 1/q_t$, where $q_t = 1/C_t$ is the marginal utility of consumption. The economy produces one type of output that can be thought of as new capital, and households can consume out of the "inventory" of capital. Because some capital is consumed, a steady state now requires that $g(m) > \delta$, rather than $g(m) < \delta$. In units of consumption, the price of capital is identically equal to 1, and so q_t is the marginal utility weighted price. Since there is no separate sector in which consumption is produced, profits from replicating capital are the only source of profits. This eliminates the term $[1 - A(k_t)]/K_t$ from (6) and so the dynamics for q_t is simply $\mathrm{D}q_t = (\rho + \delta - (g(m_t) - \mathrm{D}g(m_t)m_t))\, q_t$.

5.5.1 $g(m)/\delta \gg 1$ versus $g(m)/\delta < 1$

Suppose this economy has a steady state and let $\mathcal{E}$ be the steady state labor supply elasticity. Appendix A shows that the speed of convergence in this economy can be written as

$$\text{speed} = -\frac{\rho}{2} + \sqrt{\left(\frac{\rho}{2}\right)^2 + \frac{\mathcal{C}_g}{1-\mathcal{S}_g}\frac{g(m)}{\delta}\frac{(\rho + \mathcal{S}_g\delta)\mathcal{S}_g\delta}{\mathcal{C}_g + \frac{1-\mathcal{C}_g}{\mathcal{E}+1}}} \tag{36}$$

where $g(m) = (\rho + \delta)/(1 - \mathcal{S}_g) > \delta$. Gross output is $Kg(m)$ and investment equals δK. So $\delta/g(m)$ is the savings rate. The interest rate will be ρ and the share parameter $\mathcal{S}_g$ is the aggregate labor share. For example, $\rho = 0.04$, $\delta = 0.10$ and $\mathcal{S}_g = 0.7$ imply a savings rate equal to slightly more than 20%, and hence $g(m)/\delta \approx 5$. The factor $\mathcal{C}_g/(1 - \mathcal{S}_g)$ disappears from (36) if the technology is Cobb-Douglas. If $\mathcal{E} = \mathcal{E}_{S,v}$ and the same Cobb-Douglas technology is used in both (34) and (36), then the last factor that appears in the discriminant of (36) corresponds to $\mathcal{D}$ with $\mathcal{F} = 1$, given in (34). At $g(m)/\delta \approx 5$, (36) implies a significantly faster convergence rate in the Cass-Koopmans economy. For example, $\mathcal{E} = \mathcal{E}_{S,v} = 2$ and $\mathcal{S}_g = 1 - \mathcal{C}_g = 0.7$ in both economies implies a convergence rate of 0.125 in the economy with organization capital, and a bit more than 0.30 in the Cass-Koopmans economy. The implied half lives are 5.5 years and 2.3 years, respectively.

5.5.2 One Versus Two Types of Curvature

A striking difference between (34) and (36) is the role of the curvature parameter $\mathcal{C}_g$.

The Cass-Koopmans economy has only one curvature parameter, and the speed (36) is low when $\mathcal{C}_g$, the curvature parameter of $F(1, l) = g(l)$, is low. For example, consider a production function with a constant elasticity of substitution equal to ε, so that $\mathcal{C}_g = (1 - \mathcal{S}_g)/\varepsilon$. Such a one-sector economy has a steady state as long as $\rho + \delta$ is high enough. Furthermore, it is possible to adjust other parameters of the production function in such a way that changing ε does not change the steady-state labor share. Taking $\varepsilon \to \infty$ then makes the technology almost linear and drives the convergence rate down to zero. This slow convergence property is, of course, familiar from Jones and Manuelli [1990], where low curvature[6] generates long-run growth —that is, no convergence at all— if $\rho + \delta$ is low enough, though with completely counterfactual implications for the labor share.

The economy with organization capital has two curvature parameters that matter for the speed of convergence (in the $g(m) \uparrow \delta$ limit.) In the consumption sector, the curvature parameter of $F(1, l)$ can be written as $1 - A(1/l) + \mathcal{E}_A$, and so high curvature corresponds to a high elasticity $\mathcal{E}_A$, holding fixed the factor share of labor. This implies

[6]Note that $F(k, 1)$ and $F(1, l)$ have share and curvature parameters related by $\mathcal{S}_k\mathcal{C}_k = \mathcal{S}_l\mathcal{C}_l$ and $\mathcal{S}_l = 1 - \mathcal{S}_k$. So given share parameters, high curvature of $F(k, 1)$ is the same as high curvature of $g(l) = F(1, l)$.

relatively fast convergence—the factor $\mathcal{F}$ in (34)-(35) is large. So curvature in the consumption sector mimics the role of curvature in the Cass-Koopmans economy. But in the sector producing capital, the curvature parameter of the replication technology $G(1,m) = g(m)$ is $\mathcal{C}_g$, and so (34) says that high curvature implies slow convergence. Organization capital can be destroyed quickly but a high $\mathcal{C}_g$ limits how quickly it can be replaced.

6. A Quantitative Exploration

Some of the parameters that govern the approximate speed of convergence (34) can be identified from steady state information. In particular, firm count data point to $g(m)/\delta < 1$ close to 1 while an aggregate savings rate of 20% would imply $g(m)/\delta \approx 5$ in a Cass-Koopmans economy. But the curvature and elasticity parameters on which much depends are more difficult to infer.

In theory, the subjective discount rate ρ can be identified from the usual Euler equation $r_t = \rho + \mathrm{D}C_t/C_t$. The difficulties are numerous. Throughout the following, the conventional value $\rho = 0.04$, with time measured in years, will be used as a benchmark.

6.1 Data on Firm Counts

The steady state number of firms is $N = f(n)/\delta_f$ and the stationary distribution of organization capital across firms is Pareto on $[1, \infty)$, with a tail index equal to $\zeta = \delta_f/(g(m) - (\delta - \delta_f))$ and a mean $1/(1 - 1/\zeta)$. This implies

$$g(m) = \delta - \left(1 - \frac{1}{\zeta}\right)\delta_f, \quad f(n) = \delta_f N, \quad K = \left(1 - \frac{1}{\zeta}\right)^{-1} N. \tag{37}$$

Employment scales with organization capital, and so one can infer ζ from the tail index of the employment size distribution. In the US, this tail index is roughly $\zeta \approx 1.1$. This implies a mean capital stock per firm equal to $K/N = 11$. If capital depreciates at an overall rate of 10% and firm exit accounts for 2%, then $\delta = 0.10$ and $\delta_f = 0.02$. This implies $g(m) = 0.08 + 0.02/1.1 \approx 0.0918$. The number of employer firms in the US is roughly 6 million. On a working age population of 200 million, this implies $N = 6/200$, and thence $f(n) = 0.02 \times 6/200 = 0.0006$. Throughout everything that follows, these parameter values will be used.

The 2% entry and exit rates used here are very low compared to the 10% exit rate observed in US data, and to the 11% entry rate. Adding population growth to the model will raise the entry rate to δ_f plus the population growth rate. More importantly, post-entry firm growth in the model is completely deterministic conditional on survival. It is easy to take organization capital to be discrete and add randomness to the replication and capital depreciation processes, as in Luttmer [2011]. Then firms can exit because their last unit of capital is destroyed. As in the data, small firms then account for most of the aggregate exit rate. The formula for the tail index ζ is not affected, and the model can then account for an exit rate that is much higher than δ_f.

6.2 Data on Expenditure Shares

Write $\alpha = A(k)$ for the steady state labor share of $F(K, L)$.

6.2.1 The Savings Rate

In a steady state, the value of aggregate gross output in units of consumption is $C + \delta\widetilde{q}K$, and so the consumption rate is $c = 1/(1 + \delta qK)$. The steady state condition (22) for qK then yields

$$c = \frac{1}{1 + \frac{(1-\alpha)\delta}{\rho+\delta-(1-\mathcal{S}_g)g(m)}}. \tag{38}$$

This formula depends on the factor share parameters α and $\mathcal{S}_g$, because these parameters govern the revenues that accrue to capital, but not explicitly on $\mathcal{S}_f$. Given ρ, δ, and firm count data, (38) says that evidence on the consumption rate c imposes a linear "budget constraint" on the two factor share parameters α and $\mathcal{S}_g$. These share parameters must be in $(0, 1)$ and (38) implies that c is increasing in both. Given $\rho = 0.04$, $\delta = 0.10$ and $g(m) = 0.0918$, this implies $c \in (0.29, 1)$. This is consistent with the 20% share of US private and government investment in GDP. At $c = 0.8$, the affine relation (38) between α and $\mathcal{S}_g$ forces $\alpha \in [0.65, 0.90]$ but leaves the important share parameter $\mathcal{S}_g \in [0, 1]$ unrestricted. These bounds are essentially unchanged in the $g(m) \uparrow \delta$ limit. More information is needed to disentangle α and $\mathcal{S}_g$, and to recover $\mathcal{S}_f$.

6.2.2 Factor Shares

It will be useful to spell out the factor share implications of this economy. Workers account for a fraction $w/(\beta v + w)$ of the cost of teams producing consumption. Given a labor share α in the consumption sector, this implies worker earnings equal to

$$wL(v,w) = \frac{w}{\beta v + w} \times \alpha. \tag{39}$$

The complementary share of team factor earnings accrues to managers, but the managerial earnings $\widetilde{v}M(v,w)$ also include the income $\widetilde{v}(mK+n)$ earned by producing organization capital. Let $\ell = c[wL(v,w) + vM(v,w)]$ be the overall labor share—the ratio of the sum of worker and managerial earnings over output. These earnings are equal to $\alpha C + \widetilde{v}(mK+n)$, and therefore

$$\frac{\ell}{c} = \alpha + v(mK+n). \tag{40}$$

In turn, the first-order conditions $v = q\mathrm{D}g(m)$ and $v = q\mathrm{D}f(n)$ together with $K = f(n)/(\delta - g(m))$ and $\delta q K = (1-c)/c$ imply that the managerial services used to produce capital can be decomposed as

$$\begin{aligned} vmK &= \frac{1-c}{c}\left(\frac{g(m)}{\delta}\right)\mathcal{S}_g, & (41)\\ vn &= \frac{1-c}{c}\left(1 - \frac{g(m)}{\delta}\right)\mathcal{S}_f. & (42)\end{aligned}$$

Note that $g(m)/\delta$ is the share of aggregate investment accounted for by replication. As expected, adding up (40)-(42) shows that the overall labor share ℓ is a value-weighted average of the share parameters α, $\mathcal{S}_g$ and $\mathcal{S}_f$.

Now consider what empirical factor shares can tell us about these parameters.

Aggregate Data It is possible to combine the aggregate savings rate with the aggregate labor share to identify $\mathcal{S}_f$, but not to disentangle α and $\mathcal{S}_g$. To see this, use (41)-(42) to eliminate $v(mK+n)$ from (40) and use the resulting equation for the aggregate labor share to eliminate α from the consumption rate (38). This gives

$$\rho = \frac{1-\ell/c}{(1-c)/(\delta c)} - (\delta - g(m))\,(1-\mathcal{S}_f). \tag{43}$$

More instructively, note that $(1-c)/(\delta c) = qK$ and $(\delta - g(m))qK = qf(n)$. This means that (43) is the same as $\rho qK = 1-(\ell/c)-(1-\mathcal{S}_f)qf(n)$. That is, the value of aggregate stock of organization capital is the present value of the consumption produced, minus the cost of the workers, of the managers, and of the fixed factor used to maintain the aggregate stock of organization capital.[7] The key observation is that (43) does not depend on α or $\mathcal{S}_g$. Therefore, information about the aggregate labor share does not reveal more about α and $\mathcal{S}_g$ than is already available from (38).[8]

Backing out the fixed factor share parameter $1-\mathcal{S}_f$ from (43) and using (37) gives

$$1-\mathcal{S}_f = \frac{1}{\left(1-\frac{1}{\zeta}\right)\frac{\delta_f}{\delta}}\left(\frac{c-\ell}{1-c}-\frac{\rho}{\delta}\right).$$

Evaluated at $\zeta = 1.1$, $\delta = 0.10$ and $\delta_f = 0.02$, the leading factor on the right-hand side of this equation is equal to 55. Since $1-\mathcal{S}_f \in (0,1)$, this puts $(c-\ell)/(1-c)$ in an extremely narrow interval surrounding ρ/δ. In the basic Cass-Koopmans economy, there is no fixed factor other than labor, and it is easy to verify that this implies $(c-\ell)/(1-c) = \rho/\delta$. Here, $c = 0.8$ forces $\ell \in (0.7164, 0.7200)$. This range widens somewhat if $\delta_f/\delta \in (0,1)$ is taken to be much closer to 1 than $\delta_f/\delta = 1/5$, but not appreciably as long as ζ is as close to 1 as the approximate Zipf's law in US firm size data suggests. It is clear that plausible amounts of measurement error in c and ℓ will make any inference of $\mathcal{S}_f$ based on (43) extremely fragile.

Identifying $\mathcal{S}_f$ from aggregate data becomes impossible in the $g(m) \uparrow \delta$ limit economy described in Proposition 2. This limit economy does have a strictly positive flow $f(n)$ of entrants, but the aggregate cost of entry becomes negligible as the supply of managerial services becomes large and replication comes to dominate the accumulation of organization capital.

More Disaggregated Data Identifying α and $\mathcal{S}_g$ becomes easy with more detailed information about the sources of factor income. Together with (40), the definition of

[7] More directly, this follows from (23) together with the fact that aggregate labor income equals $w_tL(v_t,w_t) + v_tM(v_t,w_t) = \alpha + v_t(m_tK_t + n_t)$.

[8] The same is true in the two-sector economy of Uzawa [1962]. Its aggregate steady state is observationally equivalent to that of the Cass-Koopmans economy. The fact that here there is joint production does not alter this.

the overall labor share ℓ implies

$$\alpha = \left(1 - \frac{v(mK+n)}{vM(v,w) + wL(v,w)}\right)\frac{\ell}{c}. \tag{44}$$

So $\alpha \leq \ell/c$ and one can infer α from data on c and ℓ together with the fraction of all worker and managerial earnings that is earned by producing capital. In the $g(m) \uparrow \delta$ limit, (43) implies $\ell/c < 1$, and so any $v(mK+n)$ that does not exceed the earnings of managers and workers is consistent with $\alpha \in (0,1)$. Given $\rho = 0.04$, $g(m) \uparrow \delta = 0.10$, and a consumption rate $c = 0.8$, (43) implies $\ell/c = 0.9$. Data on $v(mK+n)$ together with (44) then pins down α. For example, the α implied by (44) is 0.8 if $v(mK+n)$ accounts for $1/9^{\text{th}}$ of all worker and managerial earnings. Using the $g(m) \uparrow \delta$ limit in (38) then gives $\mathcal{S}_g = ((1-\alpha)c/(1-c)) - \rho/\delta = 0.4$. More generally, the possible extremes at $c = 0.8$ and the associated $\ell/c = 0.9$ are $(\alpha, \mathcal{S}_g) = (0.9, 0)$ and $(\alpha, \mathcal{S}_g) = (0.65, 1)$, with $v(mK+n)$ varying between 0 and $5/18^{\text{th}}$ of all worker and managerial earnings.

Finally, the cost share $w/(\beta v + w)$ can be identified if it is possible to observe not only $v(mK+n)$ but also $wL(v,w)$. Combining (39) and (40) gives

$$\frac{w}{\beta v + w} = \frac{\frac{wL(v,w)}{wL(v,w)+vM(v,w)}}{1 - \frac{v(mK+n)}{vM(v,w)+wL(v,w)}}.$$

For example, if $v(mK+n)$ and $wL(v,w)$ are, respectively, $1/9^{\text{th}}$ and $3/5^{\text{th}}$ of overall worker and managerial earnings, then the cost share of workers in a team producing consumption is 67.5%.

6.3 Curvature and Elasticities

Functional form assumptions can be used to tie share parameters to curvature and elasticity parameters, but at the cost of ruling out potentially important scenarios by assumption. This section describes the quantitative implications of alternative assumptions about curvatures and elasticities.

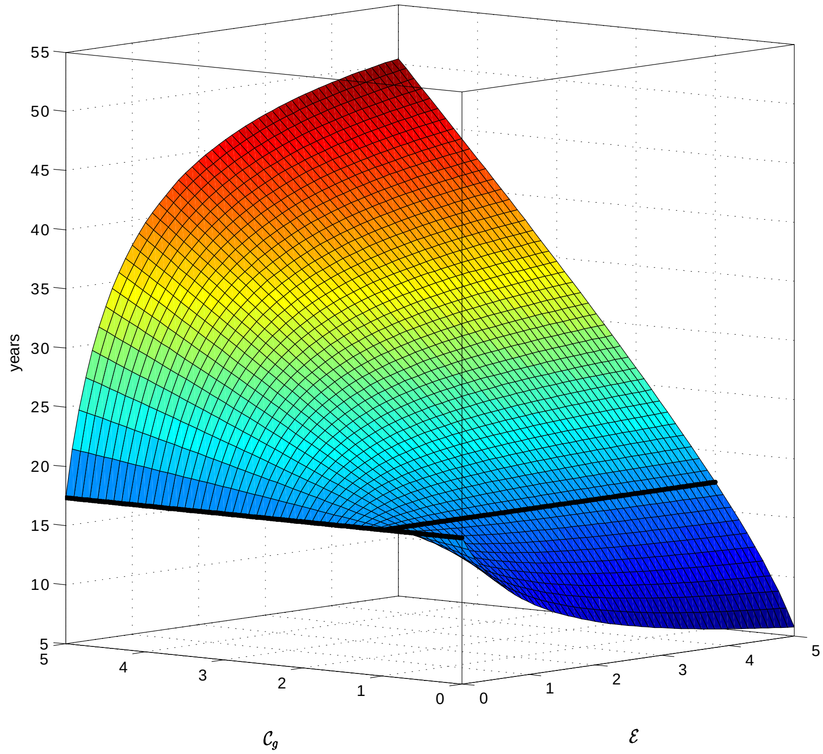

FIGURE 4 The Rate of Convergence: Half Lives Given $\mathcal{S}_g = 0.4$

To start, suppose F is Cobb-Douglas and consider the benchmark share parameters $\alpha = 0.8$ and $\mathcal{S}_g = 0.4$ that arise when $S(K,v) = v(mK+n)$ is $1/9^{\text{th}}$ of the combined worker and managerial earnings. In the special cases $(\mathcal{E}_{S,v}, \mathcal{C}_g) \in \{0\} \times (0,\infty)$ and $(\mathcal{E}_{S,v}, \mathcal{C}_g) \in [0,\infty) \times \{1\}$, the speed of convergence (26) reduces to $\delta\mathcal{S}_g$, and then $\delta = 0.10$ implies the speed $\delta\mathcal{S}_g = 0.04$. This corresponds to a half life of slightly more than 17 years, substantially longer than the 7 years associated with $\delta = 0.10$ and $\mathcal{S}_g = 1$. Figure 4 shows the half lives implied by (26), for $\rho = 0.04$, $\delta = 0.10$, $\mathcal{S}_g = 0.4$, and a range of values for $(\mathcal{E}_{S,v}, \mathcal{C}_g)$. The Cobb-Douglas case implies $\mathcal{C}_g = 1 - \mathcal{S}_g = 0.6$, and then raising $\mathcal{E}_{S,v} > 0$ automatically increases the speed of convergence relative to $\delta\mathcal{S}_g$. Figure 4 shows that speeds much lower than $\delta\mathcal{S}_g$ are possible when $\mathcal{E}_{S,v} > 0$ and $\mathcal{C}_g > 1$.

6.3.1 Factor Supply Elasticities

The speed of convergence depends on factor supply parameters only via the elasticity $\mathcal{E}_{S,v} = \mathrm{D}_v S(K,v)v/S(K,v)$ of the residual supply of managerial services. This residual supply is determined by $S(K,v) = M(v,w) - \beta L(v,w)$ and $A(K/L(v,w) = (\beta v + w)L(v,w)$, and its elasticity depends on all of the supply elasticities of $L(v,w)$ and

$M(v,w)$, as well as on the equilibrium cost share $w/(\beta v + w)$. In turn, the aggregate factor supplies $L(v,w)$ and $M(v,w)$ are the outcome of individual households solving discrete-choice problems and aggregation.[9] Everything depends on the distribution of abilities Ψ in the population: how many households are, in equilibrium, close to the margins between home production, supplying labor, and supplying managerial services?

To illustrate, simplify by assuming that $h_c = 1$ is the same for everyone. This leads to counterfactual correlations between consumption and factor supply in the cross-section, but the assumption is easy to relax. Suppose $H_A \in (0,1)$ and $H_B = 1 - H_A$, and let $\Psi(h_c, h) = H_A \Psi_A(h) + H_B \Psi_B(h)$ for all $h_c \geq 1$ and $\Psi(h_c, h) = 0$ otherwise, where Ψ_A and Ψ_B are Fréchet distributions,

$$\begin{aligned} \Psi_A(h_u, -\infty, h_w) &= \exp\left(-A_u h_u^{-\sigma} - A_w h_w^{-\sigma}\right), \\ \Psi_B(h_u, h_v, h_w) &= \exp\left(-B_u h_u^{-\theta} - B_v h_v^{-\theta} - B_w h_w^{-\theta}\right), \end{aligned}$$

for some $\sigma > 1$ and $\theta > 1$. That is, there are two populations, one that has no ability to supply managerial services, and another that does. The ability to supply managerial services might be an acquired characteristic, and then the B-population might be more educated or experienced than the A-population. The dispersion of abilities is decreasing in σ and θ, and $\sigma > 1$ and $\theta > 1$ ensures that means are finite. The distributions Ψ_A and Ψ_B give rise, of course, to logit choice probabilities (McFadden [1974]). The resulting supplies of labor and managerial services are proportional to powers, $1 - 1/\sigma$ and $1 - 1/\theta$ respectively, of these choice probabilities. The Fréchet distribution Ψ_B has the stark implication that the average earnings of type-B households are the same in both occupations. But average worker and managerial earnings need not be the same in the overall population because Ψ_A and Ψ_B differ.

Consider an economy with the following statistics

H_A	0.900
H_B	0.100
employed, fraction of population	0.700
skilled, fraction of employed	0.125
managers, fraction of employed	0.110

[9] The model is a Roy model with three activities: stay home, work, and manage. As emphasized by Heckman and Honore [1992], Roy models are hard to identify.

These statistics are enough to compute the choice probabilities for both types of households. The employment-population ratios are 68.1% and 87.5% in the type-A and type-B populations, respectively. Only 10.5% of type-B households choose to be workers, and 77% are managers.

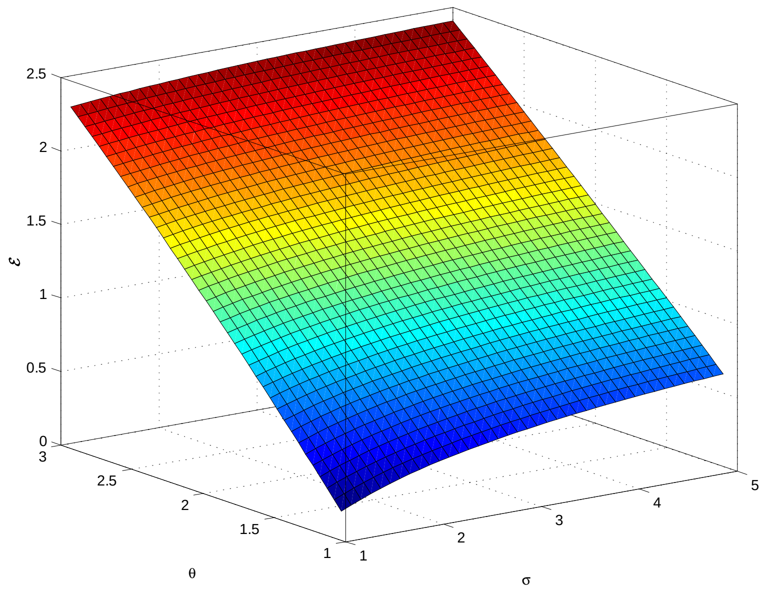

FIGURE 5 The Elasticity $\mathcal{E}$ of $S(v)$

Suppose again that F is Cobb-Douglas and consider the benchmark $c = 0.8$ with $v(mK+S)$ and $wL(v,w)$ equal to $1/9^{\text{th}}$ and $3/5^{\text{th}}$ of all worker and manager earnings. Recall that this yields $\alpha = 0.8$ and $\beta v/(\beta v + w) = 0.325$. This information suffices to infer that the mean earnings of type-A and type-B households, relative to aggregate consumption, are 0.80 and 4.68, respectively. Given a conjecture for (σ, θ), the choice probabilities and these mean earnings numbers can be used to identify $(A_u, A_w w)$ and $(B_u, B_v v, B_w w)$. In turn, this allows one to compute the equilibrium factor supply elasticities, and hence the elasticity $\mathcal{E}_{S,v}$ of $S(K,v)$. Figure 5 shows this elasticity for a range of (σ, θ). Managerial services are only supplied by type-B households. As might be expected, reducing the dispersion among type-B households makes the residual supply of managerial services more elastic. Lowering the dispersion among type-A households does so too, but not very much.

6.3.2 F Cobb-Douglas Versus CES

The conventional assumption that F is a Cobb-Douglas production function hides the important role of the labor share function $A(k)$. Suppose F is CES with an elasticity of substitution $\varepsilon \in (0,1)$. Then $A(k)$ is increasing, as required by Assumption 2.

Figure 6 shows that the factor $\mathcal{F}$ defined in (34)-(35) is a potentially important determinant of the speed of convergence. The elasticity $\mathcal{E}_A$ affects the speed not only via $\mathcal{F}$, but also because it affects the elasticities of the residual supply of managerial services. This accounts for the variation in the "Cobb-Douglas factor" $-(\rho/2) + \sqrt{(\rho/2)^2 + \mathcal{D}/\mathcal{F}}$ shown in Figure 6. Clearly, the dominant effect on the speed of an increasing $A(k)$ is through its effect on $\mathcal{F}$ itself.

Figure 7 adds variation in the main labor supply elasticity parameter σ. A more elastic supply of labor can increase the speed of convergence, but as in the Cobb-Douglas scenario displayed in Figure 5, this effect matters mostly only for very small elasticities.

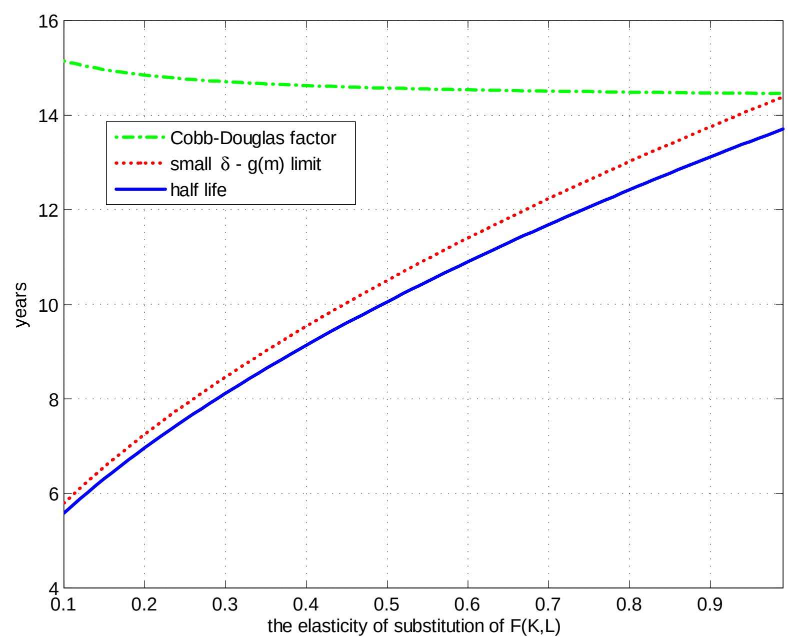

FIGURE 6 The Role of $A(k)$

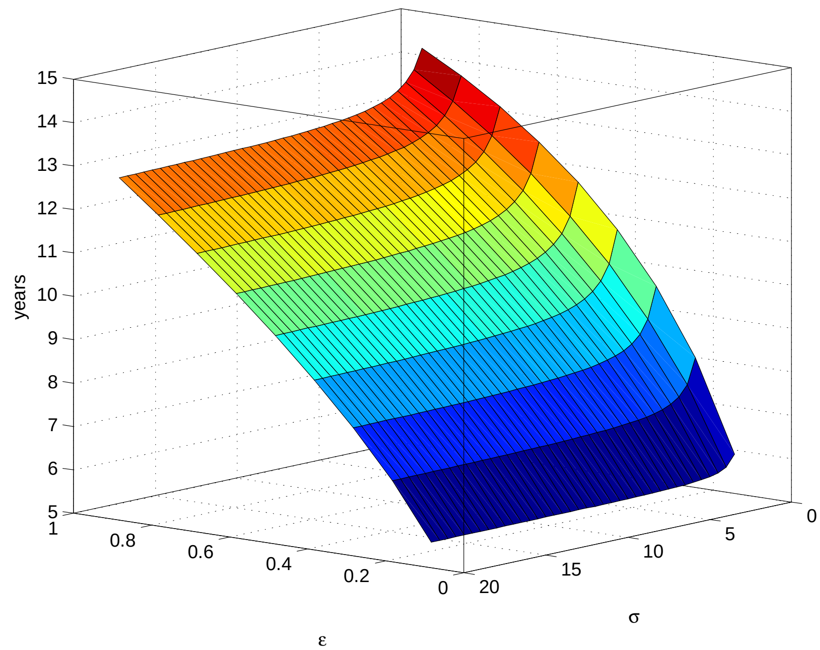

FIGURE 7 The Role $A(k)$ and the Labor Supply Elasticity Parameter σ

6.4 Equilibrium Trajectories

Figure 8 shows the CES economy with an elasticity equal to 0.5 and Cobb-Douglas versions of $g(m)$ and $f(n)$. The factor supply elasticity parameters $\sigma = 5$ and $\theta = 1.3$. This generates a half life of 10 years. Figure 9 illustrates the effect of replacing the Cobb-Douglas production function F with a CES function with an elasticity of substitution equal to $1/2$. The remaining CES parameter is chosen so that the steady state labor shares are the same for these two production functions. The half life of the Cobb-Douglas economy is 13.9 years. This falls to 10.2 years in the CES economy. Despite this, the trajectories for capital are very similar in these two economies, and both trajectories are quite close to $K_t = K + e^{-\text{speed}\times t}(K_0 - K)$. That is, the speed near the steady state remains a useful indicator for what happens over a horizon as long as 50 years.

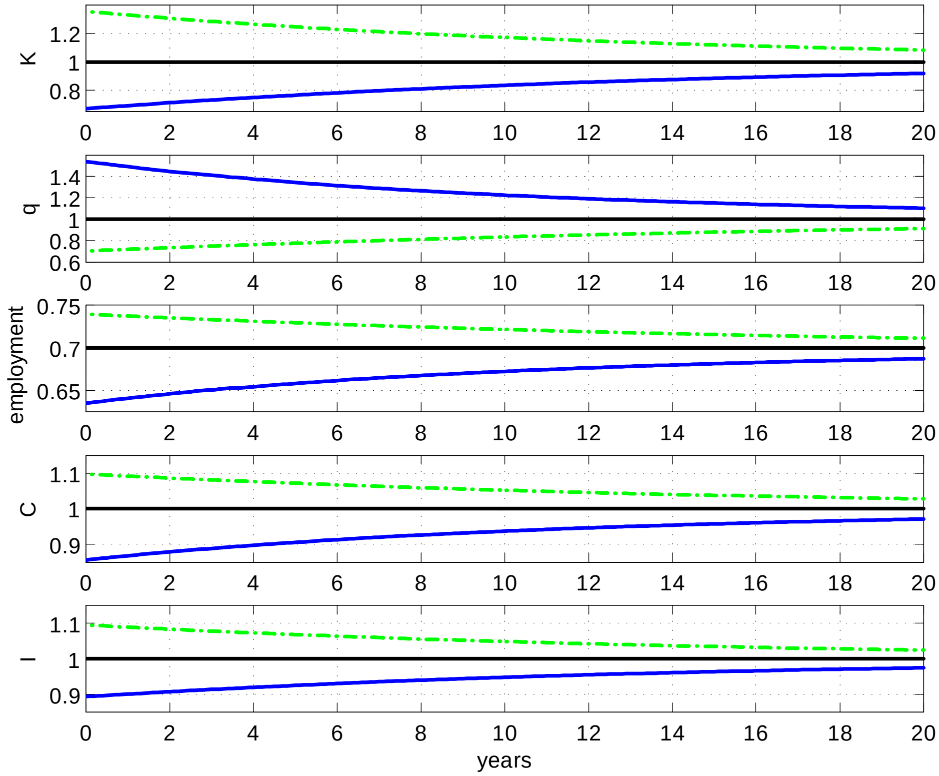

FIGURE 8 Trajectories for $K_0 < K$ and $K_0 > K$

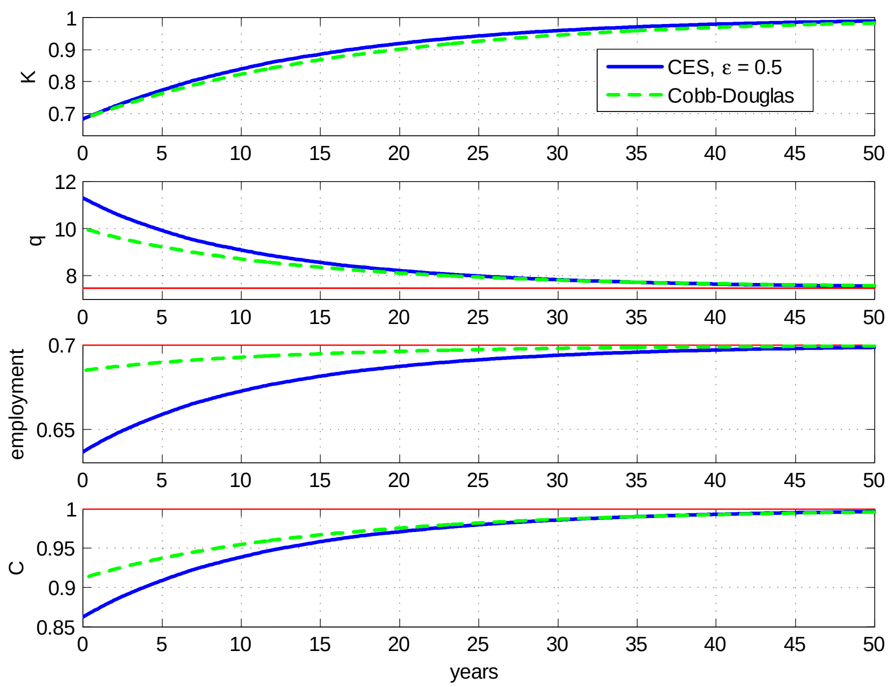

FIGURE 9 CES versus Cobb-Douglas

Although the trajectories for K_t are very similar, the trajectories for q_t, C_t, and especially for employment are not. Employment in the Cobb-Douglas economy is almost constant, as if $\beta = 0$ and the supply of labor depended only on w. On the other hand, the curvature of the CES production function is $(1 - A(k))/\varepsilon = 0.3/0.5 = 0.6$, twice what it is for the Cobb-Douglas technology. This doubling of the curvature has a strong effect on how the marginal product of labor in the consumption sector is affected by a low capital stock. Marginal-utility weighted wages at the initial date are more than 5% below the steady state, and almost 20% in units of consumption.

7. Enriching this Further

7.1 Heterogeneous Firm Growth Rates

The heterogeneity in how fast firms grow is large and persistent. As was emphasized in Luttmer [2011], this heterogeneity is precisely what is needed to account for the relatively young age of large US firms. Figure 10 provides an illustration. This section describes the implications for aggregate convergence rates.

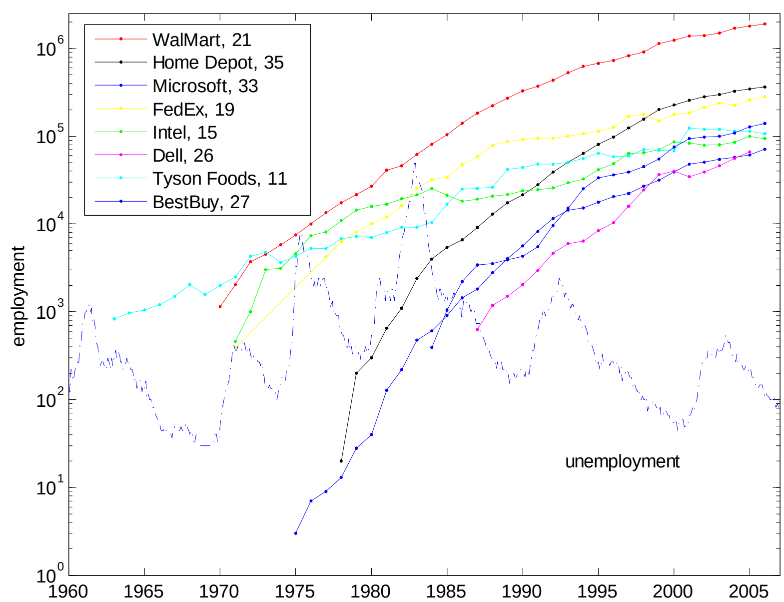

Figure 10 Rapid Firm Growth and Unemployment

To illustrate, suppose there are only two types of capital that differ in quality. High-quality capital allows workers to produce more consumption than low-quality capital. Write $K_{j,t}$ for the aggregate stock of capital of type $j \in \{\text{H,L}\}$ and $q_{j,t}$ for its marginal-utility weighted price. As before, suppose n_t units of managerial services generate units of new capital randomly at the mean rate $f(n_t)$. The type of a new unit of capital is $j \in \{\mathrm{H,L}\}$ with probability θ_j. Let $m_{j,t}$ be the managerial services used to replicate capital of type j. The technology is the same as before, for both types of capital, and so capital of type j will be replicated at the rate $g(m_{j,t})$. Finally, suppose that capital of type $j = \mathrm{H}$ changes into capital type $j = \mathrm{L}$ randomly, and once and for all, at the mean rate λ. The dynamics of the aggregate stock of capital is then

$$\mathrm{D}\begin{bmatrix} K_{\mathrm{H},t} \\ K_{\mathrm{L},t} \end{bmatrix} = -\begin{bmatrix} \delta + \lambda - g(m_{\mathrm{H},t}) & 0 \\ -\lambda & \delta - g(m_{\mathrm{L},t}) \end{bmatrix} \begin{bmatrix} K_{\mathrm{H},t} \\ K_{\mathrm{L},t} \end{bmatrix} + \begin{bmatrix} \theta_{\mathrm{H}} \\ \theta_{\mathrm{L}} \end{bmatrix} f(n_t). \tag{45}$$

The efficient rate of entry is determined by $(\theta_{\mathrm{H}} q_{\mathrm{H},t} + \theta_{\mathrm{L}} q_{\mathrm{L},t})\mathrm{D}f(n_t) = v_t$, and the optimal replication rates follow from $q_{j,t}\mathrm{D}g(m_{j,t}) = v_t$. High-quality capital will be more expensive than low-quality capital. It follows that high-quality capital will be replicated more quickly than low-quality capital. Note that $q_{\mathrm{H},t} > q_{\mathrm{L},t}$ not only because of the different profits generated by producing consumption but also because high-quality capital is replicated more quickly than low-quality capital.

If the difference between $q_{\mathrm{H},t}$ and $q_{\mathrm{L},t}$ is large near the steady state, relative to the dynamics in each of these prices near the steady state, then the gap $g(m_{\mathrm{H},t}) - g(m_{\mathrm{L},t})$ can be large even though $f(n_t)$, $g(m_{\mathrm{H},t})$ and $g(m_{\mathrm{L},t})$ are relatively stable near the steady state. Approximate this by setting $f(n_t)$ and the $g(m_{j,t})$ at their steady state values in (45). The steady state equilibrium conditions will force the eigenvalues $-(\delta + \lambda - g(m_{\mathrm{H}}))$ and $-(\delta - g(m_{\mathrm{L}}))$ to be negative. And this will be consistent with a large gap between $g(m_{\mathrm{H}})$ and $g(m_{\mathrm{L}})$ if the rate λ at which quality depreciates is high. The smallest eigenvalue, in absolute value, determines the aggregate rate of convergence.

As in the one-type case, suppose a firm is the collection of all capital that is produced by replication from a particular initial unit of capital. Let $\delta_f > 0$ be the rate which the entire firm exits, and $\delta_k = \delta - \delta_f > 0$ the rate at which individual units of its capital are destroyed. Suppose that the quality depreciation event is a firm-specific event: all units of capital in a particular firm transition from high to low quality at the same time. This implies that all units of capital within the firm are replicated at the same rate. The resulting firm size distribution is similar to the one derived for the one-quality case. In particular, suppose that $g(m_{\mathrm{H}}) - \delta_k > 0$ so that high-quality firms grow at a positive rate. Then the distribution of firm capital will behave like $k^{-\zeta}$ for large firm capital stocks k, and the tail index is given by

$$\zeta = \min\left\{\frac{\delta_f}{[g(m_\mathrm{L}) - \delta_k]^+}, \frac{\delta_f + \lambda}{g(m_\mathrm{H}) - \delta_k}\right\}. \tag{46}$$

The calculation is shown in Appendix C. US data imply that $\zeta > 1$ is close to 1. One possibility is that $g(m_\mathrm{L}) - \delta_k > 0$ and that $\zeta = \delta_f/(g(m_\mathrm{L}) - \delta_k)$ is close to 1. This means that $\delta_f + \delta_k - g(m_\mathrm{L}) = \delta - g(m_\mathrm{L}) > 0$ is close to zero. The alternative possibility is that $\zeta = (\delta_f + \lambda)/(g(m_\mathrm{H}) - \delta_k)$ is close to 1. This implies that $\delta_f + \delta_k + \lambda - g(m_\mathrm{H}) = \delta + \lambda - g(m_\mathrm{H})$ is close to zero. In either scenario, the result is that at least one of the two eigenvalues is close to zero when $\zeta > 1$ is close to 1. This implies the same connection between the aggregate convergence rate and the size distribution of firms as in the one-type economy.

But now this connection can be accounted for without invoking large firms that are much older than they are in the data (Luttmer [2011]). And this description of firm growth can account for the phenomenon illustrated by Figure 10: large firms tend to have a histories of at least several decades of very rapid growth, and these growth rates are hardly affected by the business cycle. The picture that emerges one in which employment is continuously reallocated across firms in very persistent and predictable ways, motivated by long-term considerations. These long-term considerations are the dominant determinants of firm growth, and from the perspective of rapidly growing firms, business cycles are mere blips that do not materially affect their growth trajectories.

7.2 Long-Run Aggregate Growth

It is easy to introduce exogenous economy-wide technical change that leads to long-run growth. Simply replace the technology for producing consumption by $z_t F(K_t, L_t)$ where z_t grows exogenously at some constant rate κ. Such an economy with have a steady state with a constant stock of organization capital and constant supplies of labor and managerial services. Consumption therefore grows at the same rate κ. The resulting interest rate will be $\rho + \kappa$, but the effective discount rate that governs the dynamics of (K_t, q_t) is still constant and equal to ρ, because utility is logarithmic. The marginal utility weighted prices $(q_t, v_t, w_t) = (\widetilde{q}_t, \widetilde{v}_t, \widetilde{w}_t)/C_t$ are constant in a steady state. Growth is balanced in the sense that the value of aggregate investment and of the capital stock in units of consumption grows at the same rate as consumption itself.

None of this affects the convergence rates studied in this paper. Luttmer [2012] shows how slow convergence can arise in an economy in which firm heterogeneity is driven by heterogeneity in firm productivities.

8. Concluding Remarks

Sims [1998] argued that many of the microeconomic stories underlying adjustment cost models are implausible. In the models he describes, capital accumulation amounts to opening a can of generic output, consuming some of it, and then adding the rest to the capital stock. Adjustment costs arise because the amount of output it takes is increasing and strictly convex in the rate at which capital is accumulated. It is indeed hard to tell plausible microeconomic stories for such adjustment costs. As Prescott and Visscher [1980] suggested long ago, the evidence on how firms grow is hard to interpret without thinking about the time-consuming process of accumulating some form of organization capital. It is clear that more progress can be made by constructing more explicit models of organization capital (e.g. the models of customer capital in Steindl [1965], Luttmer [2006] and Gourio and Rudanko [2014].) It would also be useful to go beyond the illustration given in Figure 10 and have more detailed evidence on how firms, and especially rapidly growing firms, respond to recessions. Much of the current evidence on how firms respond to recessions comes from repeated cross-sections, with firms typically distinguished only by size and age. This misses the persistent heterogeneity suggested by Figure 10.

The model presented here pushes the limits of what is analytically tractable in an attempt to more fully understand the features of preferences and technology that can give rise to slow convergence. It is useful to summarize the main results:

1. The fact that the firm size distribution is close to Zipf's law is a strong indication that most organization capital accumulation comes from incumbent firms expanding, and not from entry. Even if entry rates respond elastically to the state of the economy, the fact that entrants are small means that entry can do very little to speed up a recovery.

2. In the limiting economy in which Zipf's law holds, the convergence rate is increasing in the factor share of managerial services in replicating existing capital, and the

convergence rate goes to zero as this factor share goes to zero.

3. High curvature in this replication technology implies slow convergence if the supply managerial services available for replication are not completely inelastic—the marginal product of managers declines too quickly for it to be efficient to allocate large amounts of managerial resources to replication. This effect is magnified by the particularly elastic residual supply of managerial services that arises when the managerial share in the cost of teams of managers and workers is substantial.

4. As in the one-sector Cass-Koopmans economy, high curvature in the technology for producing consumption implies relatively fast convergence: with high curvature, a low capital stock implies a high capital share, and this adds to the incentives to produce more capital.

5. For a robust set of parameters, aggregate consumption, investment, and employment are all below their steady states when the stock of organization capital is below its steady state.

The analysis in this paper is motivated by the belief that rather than thinking of recessions as the result of negative productivity shocks, it more useful to think of recessions as events in which organization capital is destroyed. Negative productivity shocks have been derided as periods of collective amnesia, and with reason. But the ways in which organization capital can be destroyed are legion. For example, Luttmer [2013] describes a related economy in which a news shock causes low-quality capital to be abandoned when high-quality capital is found to be below its steady state.

Households in the economy described in this paper are heterogeneous in terms of their ability to supply labor and managerial services. But these abilities are taken to be fixed and general—they are not acquired over time or specific to particular occupations or particular industries or firms. Adding this type of heterogeneity will complement the sources of slow convergence highlighted in this paper.

References

[1] Beaudry, and Portier, "When Can Changes in Expectations Cause Business Cycle Fluctuations in Neo-Classical Settings?," vol. 135, issue 1 (2007), 458-477.

[2] Brock, W.A. and J.A. Scheinkman, "Global Asymptotic Stability of Optimal Control Systems with Applications to the Theory of Economic Growth," *Journal of Economic Theory*, vol. 12 (1976), 164-190.

[3] Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll, "The Dynamics of Inequality," *Econometrica*, vol. 84, no. 6 (2016), 2071–2111.

[4] Gourio, F. and L. Rudanko, "Customer Capital," *Review of Economic Studies*, vol. 81, no. 3, (2014), 1102-1136.

[5] Hayashi, F., "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, vol. 50, no.1 (1982), 213-224.

[6] Honoré, B.E. and J.J. Heckman, "The Empirical Content of the Roy Model," *Econometrica*, vol. 58, no. 5 (1990), 1121-1149.

[7] Jones, L.E. and R. Manuelli, "A Convex Model of Equilibrium Growth: Theory and Policy Implications," *Journal of Political Economy*, vol. 98, no. 5 (1990), 1008-1038.

[8] Lucas, R.E., Jr., "Adjustment Costs and the Theory of Supply," *Journal of Political Economy*, vol. 75, no. 4 (1967), 321-334.

[9] Luttmer, E.G.J., "Consumer Search and Firm Growth," *Federal Reserve Bank of Minneapolis working paper 645* (2006).

[10] Luttmer, E.G.J., "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, vol. 2 (2010).

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

[12] Luttmer, E.G.J., "Slow Convergence in Economies with Firm Heterogeneity," *Federal Reserve Bank of Minneapolis working paper 696* (2012).

[13] Luttmer, E.G.J., "The Stolper-Samuelson Effect of a Decline in Aggregate Consumption," *Federal Reserve Bank of Minneapolis working paper 703* (2013).

[14] McFadden, D.L., "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), Frontiers of Econometrics, Academic Press (1974).

[15] Prescott, E.C. and M. Visscher, "Organization Capital," *Journal of Political Economy*, vol. 88, no. 3 (1980), 446-461.

[16] Rebelo, S., "Long-Run Policy Analysis and Long-Run Growth," *Journal of Political Economy*, vol. 99, no. 3 (1991), 500-521.

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

[18] Sims, C.A., "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, vol. 49 (1998), 317-356.

[19] Steindl, J., *Random Processes and the Growth of Firms: A Study of the Pareto Law*, Hafner (1965).

[20] Uzawa, H., "On a Two-Sector Model of Economic Growth," *Review of Economic Studies*, vol. 29, no. 1 (1961), 40-47.