Research Division
FEDERAL RESERVE BANK *of* MINNEAPOLIS

Slow Convergence in Economies with Organization Capital

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

Working Paper 748
January 2018

DOI: https://doi.org/10.21034/wp.748
Keywords: Business cycles; Firm size distribution; Slow recoveries; Zipf's law
JEL classification: E32, L11

Federal Reserve Bank of Minneapolis • 90 Hennepin Avenue • Minneapolis, MN 55480-0291
https://www.minneapolisfed.org/research/

Slow Convergence in Economies with Organization Capital*

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

January 19, 2018

Abstract

Most firms begin very small, and large firms are the result of typically decades of persistent growth. This growth 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 aggregate convergence rates that are only about half of what they are in the standard Cass-Koopmans economy. 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

Between December 2007 and December 2009, the US civilian employment-population ratio plunged from 62.7% to 58.3%. As of December 2017, it has recovered to only 60.1%.

*This is a substantially revised and expanded version of my working paper "Slow Convergence in Economies with Firm Heterogeneity" (Luttmer [2012]). More detail on some of the proofs is available at `www.luttmer.org`.

If the gap between trough and peak is shrinking at a constant rate, this implies a half-life of about $8 \times \ln(4.4/2.6)/\ln(2) \approx 6.1$ years. Why is it taking so long for employment to recover? This paper argues that such a slow recovery is a robust implication of a model of the aggregate economy with heterogeneous firms that grow by accumulating some sort of organization capital.

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 (the $\zeta \downarrow 1$ limit is known as Zipf's law.) Most new firms start out with only a few employees, and it took the largest firms in the US economy decades of rapid and persistent growth to reach their current size.[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 exit randomly 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 particular, Zipf's law implies no aggregate convergence at all. In the US, firm entry and exit rates are around 10% per annum, and so the implied aggregate mean reversion rate is $\varepsilon - g \approx (1 - 1/1.1)0.1 \approx 0.01$. This implies a half-life of almost 70 years. Even longer half-lives emerge when not all exit is random (Luttmer [2011]).

Obviously, the US economy recovers more quickly from recessions than suggested by this simple calculation. But this account of firm growth and aggregate convergence conveys an important intuition: entry, exit, and non-stationarity at the level of individual firms very naturally add up to stationary thick-tailed cross-sectional distributions and slow aggregate convergence.[2] This paper builds on this observation by interpreting firm

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

[2] Not unlike in Penrose [1959], surviving firms in this economy keep growing. This is in contrast to,

growth as organization capital accumulation and allowing the flow of entrants f and the rate g at which firms choose to accumulate organization capital to respond to the state of the economy. The convergence rate of the economy is then characterized in terms of factor supply elasticities, factor share parameters, and curvature parameters. This characterization continues to suggest an important connection between slow aggregate convergence and skewed firm size distributions. A benchmark calibration, in which the quality of organization capital is heterogeneous across firms, so that different types of firms choose to grow at different rates, produces a half-life of slightly more than 6 years.

An attempt to argue that the destruction of organization capital (or, alternatively, a sudden realization that the stock of organization capital is below its long-run steady state) can be an important element of what happens in recessions has to overcome the implied wealth effect on labor supply. In a Cass-Koopmans economy with an elastic supply of labor and standard preferences, a low capital stock tends to result in high investment that is made possible by a combination of both low consumption and high employment. High employment is the opposite of what characterizes a typical recession. This difficulty is avoided here by introducing managers and workers, and recognizing that they are not used with the same intensity in producing consumption and new organization capital. Workers are assumed to supply labor that can only be used to produce consumption goods, under the supervision of managers, while managers can divide their time between two distinct tasks: overseeing workers and producing new organization capital. The need for managerial supervision means that fewer workers can be employed when managers are particularly busy producing new organization capital.[3]

In the model, managers can produce new organization capital by replicating existing organization capital, which leads to incumbent firm growth, or by using a fixed factor (scientists, certain locations, say) to create new organization capital from scratch, interpreted as entry. The equilibrium g and f will be above their steady state levels when the organization capital stock, worker employment, and aggregate consumption are below their steady states. This endogenous response of organization capital accumulation to the level of the capital stock ensures that, even in the $\zeta \downarrow 1$ limit, the economy recovers

for example, Atkeson and Kehoe [2005] and much of the large literature that follows the tradition of Hopenhayn [1992] and Hopenhayn and Rogerson [1993].

[3]This is taken from Luttmer [2013]. Many authors use the Greenwood, Hercowitz, and Huffman [1988] preference specification to eliminate the wealth effect on labor supply.

at a strictly positive rate. The Zipf limit $\zeta \downarrow 1$ arises when labor and managerial services inputs are abundant relative to the fixed factor required for firm entry. Although entry rates remain positive in this limit, the contribution of entry to new employment becomes negligible relative to the contribution of incumbent firm growth. The Zipf limit produces a good approximation for the observed firm size distribution, and the rate of convergence of the economy varies continuously as $\zeta \downarrow 1$. The question then arises: what governs the speed of adjustment of the $\zeta \downarrow 1$ economy?

The paper gives detailed analytical answers to this question by making two strong parametric assumptions. One is the fairly common simplifying assumption that flow utility is a logarithmic function of consumption and additively separable across consumption and leisure. The second assumption is that consumption is produced using organization capital and team services, with a technology for team services that is Leontief in labor and managerial services. Everything else is non-parametric. In particular, there are separate roles for factor share parameters and curvature parameters (or, implicitly, substitution elasticities).

Throughout, an important determinant of the speed of convergence is the factor share of managers in replicating organization capital. A small factor share means that g cannot vary much with the supply of managerial services per unit of capital, and hence, other things equal, with the aggregate capital stock itself. This leads to slow convergence. Slow convergence also follows when the curvature of g, as a function of managerial services per unit of capital, is high. Strongly diminishing returns then imply that it cannot be optimal to increase managerial services per unit of capital much beyond what is optimal in the steady state.

The elasticity of substitution between capital and team services in the production of consumption plays an important role. If this elasticity is below one, then the factor share of capital is increasing in the ratio of team services to capital (and thus the labor-capital ratio, given that the technology for team services is Leontief). This implies a high capital share when the capital stock is low, creating strong incentives to produce more capital. A consumption sector with a low elasticity of substitution between capital and team services therefore speeds up the rate at which the economy converges.[4] The

[4] This is also true in the Cass-Koopmans economy. Jones and Manuelli [1990] produce long-run growth (no convergence) by assuming a very high elasticity of substitution.

effect of this on the shape of equilibrium trajectories can be quite strong. In an economy that operates below its steady state, the incentives to reallocate managers away from overseeing workers are weak when the technology is Cobb-Douglas, and strong enough to create a sharp drop in worker employment when the elasticity of substitution between capital and team services is significantly below one.

In an economy with two types of capital, it may well be that only firms with high-quality capital choose to replicate organization capital. In the benchmark calibration that generates a half-life of slightly over 6 years, the number of such firms is relatively small, and they grow at an annual rate of about 25% in the steady state. Low-quality firms choose not to grow at all. Furthermore, the curvature in g is strong enough that fast-growing firms choose not to grow much faster when the economy is not too far below its steady state. Figure 1 in Luttmer [2012] provides suggestive evidence that this is indeed the case. At the same time, the higher capital prices that prevail when the economy is below its steady state do not pull low-quality firms off their corner: organizational capital may be a bit more valuable, but not by enough to turn a stagnant firm into a "gazelle."[5]

The picture that emerges is one in which there is a continuous reallocation of labor from low-quality firms to high-quality firms, both in the steady state and away from it. When a recession hits, say because some low-quality organization capital is destroyed, fast-growing firms do speed up somewhat, but not by enough to quickly absorb the workers fired by the firms that experienced the destruction of low-quality capital.

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 pointed out in Luttmer [2012]. The idea is reminiscent of how Granger [1980] obtained long memory processes, by aggregating heterogeneous but still stationary processes. 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 argues that the equilibrium trajectories associated with slow aggregate convergence are not dissimilar to what has happened following recent US recessions. The observation that recent recessions seem to give rise to

[5] The persistent rapid growth of some firms led Birch [1979] to introduce the term "gazelles" to describe such firms. There are also many small firms that hardly grow (Hurst and Pugsley [2011]).

slow (some argue “jobless”) recoveries has received significant attention.[6] In particular, organization capital is a key factor in the account of Koenders and Rogerson [2005]. But the underlying firm dynamics in their model is too simple to make contact with the evidence on how firms grow and the resulting thick-tailed stationary size distributions.

Outline Section 2 lays out the economy in detail. Section 3 describes the steady state implications for the firm size distribution. The circumstances in which this size distribution approximates Zipf’s law are given in Section 4. The characterization of the speed of convergence is in Section 5. Section 6 introduces heterogeneity in the quality of organization capital and describes quantitatively how the model can generate recessions followed by slow recoveries.

2. The Economy

Organization capital is taken to be a type of capital that can be used simultaneously to produce consumption and more organization capital. But the technologies for producing consumption and new organization capital are assumed to be different. Labor is a specific factor for the consumption sector, while managerial services are used in both sectors. Heterogeneous ability and comparative advantage determine the aggregate supplies of labor and managerial services.

2.1 Households

There is 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 assumed to be time invariant and denoted by Ψ. Time is continuous, and household preferences are recursive and additively separable across time. The contribution of a type-h household member to flow utility at time t is $h_c \ln(C_t(h)) + h_u(1 - \iota_t(h))$, where $C_t(h)$ is flow consumption and $\iota_t(h) \in \{0, 1\}$ is a labor market participation decision. It will be convenient to normalize the mean of h_c to be equal to 1, so that the marginal utility of household consumption is going to be $1/C_t$ when

[6] Notable examples are Bachmann [2012], Berger [2016], Jaimovich and Siu [2015] and Koenders and Rogerson [2005].

$C_t = \int C_t(h)\Psi(\mathrm{d}h)$. The distribution Ψ is also assumed to be sufficiently smooth that the employment lotteries proposed in Rogerson [1988] are not needed. Household preferences over flows of consumption, labor, and managerial services are then

$$\mathcal{U}(C, L, M) = \int_0^\infty e^{-\rho t} \left(\ln(C_t) - V(L_t, M_t)\right) \mathrm{d}t,$$

where ρ is strictly positive, and where $V(L, M)$ is the household disutility from supplying L units of labor and M units of managerial services,

$$\begin{aligned} V(L, M) \quad = \quad & \min_{\substack{(\iota_v(\cdot), \iota_w(\cdot)) \in \{0,1\}^2 \\ \iota_v(\cdot) + \iota_w(\cdot) \leq 1}} \int h_u[\iota_v(h) + \iota_w(h)]\Psi(\mathrm{d}h) \\ & \text{s.t.} \int h_w \iota_w(h)\Psi(\mathrm{d}h) \geq L, \quad \int h_v \iota_v(h)\Psi(\mathrm{d}h) \geq M. \end{aligned}$$

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 the 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. 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. 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. 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

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

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

The fact that these supply curves only depend on the marginal utility weighted factor prices 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—a potentially important complication that is abstracted from here. The following assumption and lemma summarize the important properties of (1)-(2).

Assumption 1 *The type distribution Ψ has finite mean and is sufficiently smooth to ensure that $V(L, M)$ is continuously differentiable.*

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

$$\mathrm{D}_1 M(v, w) \geq 0, \quad \mathrm{D}_2 L(v, w) \geq 0, \quad \mathrm{D}_2 M(v, w) = \mathrm{D}_1 L(v, w) \leq 0.$$

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

The symmetry follows because $L_t = L(v_t, w_t)$ and $M_t = M(v_t, w_t)$ solve $[w_t, v_t] = \mathrm{D}V(L_t, M_t)$, and $\mathrm{D}^2 V(L_t, M_t)$ will be symmetric. 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 curves, evaluated at equilibrium prices and quantities. 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 $(0, 1)$.

These aggregate factor supply elasticities abstract from effort and 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 low h_w/h_u.

2.2 The Technology for Producing Consumption

Consumption is produced using capital K_t and the services of a team of managers and workers. The technology for team services is Leontief. The input requirements for a unit of team services are one unit of labor and β units of managerial 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 constant returns to scale production function that is assumed to be smooth and strictly increasing in both factors. Because of the logarithmic utility assumption, the production function F will turn out to affect the dynamic properties of this economy only via the factor share

$$A(k) = \frac{\mathrm{D}_2 F(k,1)}{F(k,1)}. \tag{3}$$

This is the factor share of 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

$$(\beta v_t + w_t) L(v_t, w_t) = A(k_t), \quad k_t = \frac{K_t}{L(v_t, w_t)}. \tag{4}$$

This determines functions $w_t = w(K_t, v_t)$ and $k_t = k(K_t, v_t)$ that describe how wages and the capital-labor ratio vary with K_t and v_t in any equilibrium.[7] 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.[8]

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 increasing 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

[7] 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.

[8] Section 6 generalizes by introducing heterogeneity in the quality of capital. There, an increasing $A(\cdot)$ implies that the factor share of high-quality capital is also high.

then the dependence of the right-hand side of (4) on K_t vanishes. The capital stock can still affect wages and worker employment in that case, but only indirectly through its effect on 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 it lowers the supply of labor because marginal households switch from supplying labor to supplying managerial services. Only this second channel remains if $\beta = 0$, and then (4) 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 supply of labor are both constant. Lemma 2 describes the much richer set of possibilities that can arise when β is positive and F is not Cobb-Douglas.

Lemma 2 *Write $\mathcal{E}_{A,t}$ for the elasticity of the labor share function $A(\cdot)$ defined in (3). Let $w_t = w(K_t, v_t)$ and $k_t = k(K_t, v_t)$ be the solution to (4). Assumptions 1 and 2 imply that these functions are well defined, and their elasticities are*

$$\begin{bmatrix} \mathcal{E}_{w,K,t} & \mathcal{E}_{w,v,t} \\ \mathcal{E}_{k,K,t} & \mathcal{E}_{k,v,t} \end{bmatrix} = \frac{1}{\frac{w_t}{\beta v_t + w_t} + (1 + \mathcal{E}_{A,t})\, \mathcal{E}_{L,w,t}} \times \begin{bmatrix} \mathcal{E}_{A,t} & -\left(\frac{\beta v_t}{\beta v_t + w_t} + (1 + \mathcal{E}_{A,t})\, \mathcal{E}_{L,v,t} \right) \\ \frac{w_t}{\beta v_t + w_t} + \mathcal{E}_{L,w,t} & \left(\frac{\beta v_t}{\beta v_t + w_t} \right) \mathcal{E}_{L,w,t} - \left(\frac{w_t}{\beta v_t + w_t} \right) \mathcal{E}_{L,v,t} \end{bmatrix}.$$

This implies $\mathcal{E}_{k,K,t} \in (0, 1]$ and $\mathcal{E}_{w,K,t} \geq 0$. And $\mathcal{E}_{k,v,t} > 0$ but the sign of $\mathcal{E}_{w,v,t} < -\mathcal{E}_{L,v,t} / \mathcal{E}_{L,w,t} < 1$ is ambiguous.

The elasticity $\mathcal{E}_{w,v,t}$ is strictly negative if the cross price elasticities $\mathcal{E}_{L,v,t}$ and $\mathcal{E}_{M,w,t}$ are zero. More generally, the sign of $\mathcal{E}_{w,v,t}$ is ambiguous because the expenditures $(\beta v_t + w_t) L(v_t, w_t)$ on teams of managers and workers may rise or fall with v_t. Holding fixed K_t, the factor prices v_t and w_t will move in opposite directions if the factor share of managers in teams of managers and workers is substantial.

2.2.1 The Residual Supply of Managerial Services

Using the wage function $w_t = w(K_t, v_t)$ implied by (4), define

$$S(K_t, v_t) = M(v_t, w(K_t, v_t)) - \beta L(v_t, w(K_t, v_t)). \tag{5}$$

This is the residual supply of managerial services available to the sector producing new capital. The elasticities $\mathcal{E}_{S,K,t}$ and $\mathcal{E}_{S,v,t}$ of $S(K_t, v_t)$ are easy to compute from (5) and

Lemma 2. 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. More generally, assuming for the simplicity of this argument that all elasticities are non-zero, the upper bound $\mathcal{E}_{w,v,t} < -\mathcal{E}_{L,v,t}/\mathcal{E}_{L,w,t}$ together with (5) implies

$$\frac{\mathcal{E}_{S,v,t}}{\mathcal{E}_{M,v,t}} > \frac{M(v_t, w_t)}{S(K_t, v_t)} \left(1 - \frac{\mathcal{E}_{M,w,t}}{\mathcal{E}_{M,v,t}} \times \frac{\mathcal{E}_{L,v,t}}{\mathcal{E}_{L,w,t}}\right).$$

This lower bound is positive because cross price elasticities are smaller than own price elasticities.

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.*

2.3 The Technology for Producing New Capital

New capital can be produced in two ways. 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. Or there could be a fixed supply of experts whose inputs are needed for every start-up.

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

$$\mathrm{D}K_t = (g(m_t) - \delta)K_t + f(n_t). \tag{6}$$

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

This economy features joint production: 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 is the capital-labor ratio determined by (4). That is, the required return on a unit of capital comes in the form of earnings from producing consumption, earnings from replicating capital, and 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 $\widetilde{v}_t = \widetilde{q}_t \mathrm{D}f(n_t)$. This is where Assumption 3 comes in to avoid the need to consider corners. Write $q_t = \widetilde{q}_t / C_t$ for the marginal utility weighted price of a unit of capital. The first-order conditions for m_t and n_t are then

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

and the Euler condition $r_t = \rho + \mathrm{D}C_t / C_t$ yields

$$\rho q_t = \frac{1 - A(k_t)}{K_t} + q_t(g(m_t) - \delta) - v_t m_t + \mathrm{D}q_t. \tag{7}$$

As is standard, the optimality conditions for this economy also include a transversality condition that requires $e^{-\rho t} q_t K_t$ to go to zero as t becomes large.

Share and Curvature Parameters It is convenient to write $\mathcal{S}_{g,t} = \mathrm{D}g(m_t)m_t/g(m_t)$ for the managerial factor share and $\mathcal{C}_{g,t} = -\mathrm{D}^2 g(m_t)m_t/\mathrm{D}g(m_t)$ for the curvature of $g(\cdot)$, evaluated at the equilibrium value of m_t. The implied elasticity of substitution between managerial services and organization capital is $(1 - \mathcal{S}_{g,t})/\mathcal{C}_{g,t}$. The analogous share and curvature parameters for $f(\cdot)$ are $\mathcal{S}_{f,t}$ and $\mathcal{C}_{f,t}$.

2.4 The Static Equilibrium Conditions

Holding fixed K_t and q_t, the static equilibrium conditions for this economy can be summarized as

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

This requires solving $[\mathrm{D}g]^{-1}(v_t/q_t)K_t + [\mathrm{D}f]^{-1}(v_t/q_t) = S(K_t, v_t)$ for v_t and then m_t and n_t follow. The fact that $S(K_t, \cdot)$ is upward sloping and Assumption 3 imply that the solution is unique and interior. Inspection of the static equilibrium conditions (8) shows that m_t and n_t must co-move. More precisely, the first-order conditions $v_t = q_t \mathrm{D}g(m_t)$ and $v_t = q_t \mathrm{D}f(n_t)$ imply

$$\begin{bmatrix} \frac{K_t}{m_t}\frac{\partial m_t}{\partial K_t} & \frac{q_t}{m_t}\frac{\partial m_t}{\partial q_t} \\ \frac{K_t}{n_t}\frac{\partial n_t}{\partial K_t} & \frac{q_t}{n_t}\frac{\partial n_t}{\partial q_t} \end{bmatrix} = \begin{bmatrix} 1/\mathcal{C}_{g,t} \\ 1/\mathcal{C}_{f,t} \end{bmatrix} \begin{bmatrix} -\frac{K_t}{v_t}\frac{\partial v_t}{\partial K_t} & 1 - \frac{q_t}{v_t}\frac{\partial v_t}{\partial q_t} \end{bmatrix}, \tag{9}$$

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).[9] The relative magnitude of the elasticities of m_t and n_t depends only on the curvatures of $g(\cdot)$ and $f(\cdot)$: m_t responds less than n_t if and only 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) implied by (8) gives

$$\begin{bmatrix} \frac{K_t}{v_t}\frac{\partial v_t}{\partial K_t}, & \frac{q_t}{v_t}\frac{\partial v_t}{\partial q_t} \end{bmatrix} = \frac{\begin{bmatrix} \frac{m_t K_t}{m_t K_t + n_t} - \mathcal{E}_{S,K,t}, & \frac{m_t K_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{g,t}} + \frac{n_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{f,t}} \end{bmatrix}}{\mathcal{E}_{S,v,t} + \frac{m_t K_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{g,t}} + \frac{n_t}{m_t K_t + n_t}\frac{1}{\mathcal{C}_{f,t}}}. \tag{10}$$

From Lemma 3, 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

[9] All prices and quantities at a point in time are functions only of (K_t, q_t). Here and below, the notation $\partial/\partial K_t$ and $\partial/\partial q_t$ will be used exclusively for the partial derivatives of these functions.

to q_t is in $(0,1)$. The intuition is that an increase in q_t is distributed over an increase in the factor 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 (9) and (10) shows that m_t and n_t are both decreasing in K_t and increasing in q_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 dependence on (K_t, q_t) of the capital-labor ratio implied by (4) is characterized by

$$\begin{bmatrix} \frac{K_t}{k_t}\frac{\partial k_t}{\partial K_t} & \frac{q_t}{k_t}\frac{\partial k_t}{\partial q_t} \end{bmatrix} = \begin{bmatrix} \mathcal{E}_{k,K,t} + \mathcal{E}_{k,v,t} \times \frac{K_t}{v_t}\frac{\partial v_t}{\partial K_t} & \mathcal{E}_{k,v,t} \times \frac{q_t}{v_t}\frac{\partial v_t}{\partial q_t} \end{bmatrix}, \tag{11}$$

where $\mathcal{E}_{k,K,t}$ and $\mathcal{E}_{k,v,t}$ are defined in Lemma 2. Together with (10) this implies that the capital-labor ratio in the consumption sector is increasing in both K_t and q_t.

2.5 Equilibrium

The derivatives $\mathrm{D}K_t$ and $\mathrm{D}q_t$ given in (6) and (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 (4) 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. As usual, the transversality condition will be needed to pin down q_0.

2.6 Alternative Formulations

The notion of "organization capital" adopted here is abstract. It can be made more explicit by taking capital to be discrete at the micro level. One unit of capital could then be a blueprint that can be used at the same time by only one team or a few teams of managers and workers, or in only a restricted number of geographical locations. And $g(\cdot)$ can then be interpreted as a Poisson arrival rate that describes the rate at which blueprints can be replicated. Models of customer capital (e.g., Steindl [1965], Luttmer

[2006], and Gourio and Rudanko [2014]) have a very similar structure.

The Trouble with Homogeneous Labor Inputs The properties of the residual supply curve of managerial services $S(K_t, v_t)$ described in Lemma 3 continue to hold, and are easy to see by inspection, if it is assumed that labor and managerial services are perfect substitutes. Given an appropriate choice of units, $w_t = v_t$ at all times, and (4) becomes an equilibrium condition that determines a labor demand curve $\mathcal{L}(K, v)$ via $(1 + \beta)v\mathcal{L}(K, v) = A(K/\mathcal{L}(K, v))$. It is easy to see that Assumption 2 implies that $\mathcal{L}(K, v)$ is increasing in K and decreasing in v. Assumption 2 also implies that the implied capital-labor ratio $k = K/\mathcal{L}(K, v)$ is increasing in both K and v. The residual supply of labor not used to produce consumption is then $S(K, v) = M(v, v) + L(v, v) - (1 + \beta)\mathcal{L}(K, v)$. This is clearly weakly decreasing in K and Lemma 1 ensures that it is strictly increasing in v.

The cost of this simplification is that it forces both v_t and w_t and $M(v_t, w_t)$ and $L(v_t, w_t)$ to move together along any equilibrium trajectory. This rules out the possibility that different types of labor are affected differently by a destruction of organization capital. In particular, if F is Cobb-Douglas, then $\mathcal{L}(K, v)$ no longer depends on K and this would rule out the possibility that a destruction of organization capital could cause a reduction in aggregate employment.

Long-Run Growth A simple way to introduce long-run growth is to replace $C_t = F(K_t, L_t)$ by $C_t = z_t F(K_t, L_t)$ with z_t growing exponentially. The formulation of preferences ensures that the supplies of labor and managerial services are constant when consumption and factor prices grow at a common rate. Because z_t only affects the output of consumption goods, growth is balanced even though production functions need not be Cobb-Douglas. Much richer formulations, in which capital quality is heterogeneous and growth is endogenous, are possible but beyond the scope of the current paper.

Monopolistic Competition It is possible to re-interpret K_t as the number of goods in an economy with monopolistic competition and a technology for producing differentiated commodities that is linear in team services. If C_t is a symmetric CES composite good of differentiated commodities, with an elasticity of substitution greater than 1, then the

$A(\cdot)$ in (4) is a constant equal to 1 minus the reciprocal of the elasticity of substitution. The resulting economy is isomorphic to a competitive economy in which F is Cobb-Douglas. Everything that follows for the competitive Cobb-Douglas economy applies. For more general but still symmetric preferences over differentiated commodities, the elasticity of the demand curves faced by individual producers will depend only on the number of goods K_t. The function $A(\cdot)$ in (4) is then no longer a function of the capital-labor ratio $K_t/L(v_t, w_t)$, but of the number of goods K_t only. This makes a difference for the dynamic properties of this economy. A detailed analysis is left to future work.

3. 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 the same firm. From (6), 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 results in 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, these inequalities imply 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 both in 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 (6), 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.

3.1 An Economy with Two Types of Capital

Growth rates have an enormous amount of persistent heterogeneity that any calibration will have to confront. Without exception, very large firms have had histories of very rapid and persistent growth, and large firms account for an important share of aggregate employment. As was emphasized in Luttmer [2011], this heterogeneity is precisely what is needed to account for the relatively young age of large US firms.

To show how this heterogeneity affects the connection between aggregate convergence rates and the firm size distribution, suppose there are only two types of capital that differ in quality.[10] 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 $j \in \{1, 2\}$ 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 $j \in \{1,2\}$ with probability ϕ_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 $j = 2$ changes into capital type $j = 1$ randomly, and once and for all, at the mean rate θ. The dynamics of the aggregate stock of capital is then

$$\mathrm{D}\begin{bmatrix} K_{1,t} \\ K_{2,t} \end{bmatrix} = -\begin{bmatrix} \delta - g(m_{1,t}) & -\theta \\ 0 & \delta + \theta - g(m_{2,t}) \end{bmatrix}\begin{bmatrix} K_{1,t} \\ K_{2,t} \end{bmatrix} + \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} f(n_t). \tag{12}$$

The efficient rate of entry is determined by $(\phi_1 q_{1,t} + \phi_2 q_{2,t})\mathrm{D}f(n_t) = v_t$, 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 $q_{2,t} > q_{1,t}$ 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 $q_{2,t}$ and $q_{1,t}$ is large near the steady state, relative to the dynamics in each of these prices near the steady state, then the gap $g(m_{2,t}) - g(m_{1,t})$ can be large even though $f(n_t)$, $g(m_{1,t})$, and $g(m_{2,t})$ are relatively stable near the steady state. Approximate this by setting $f(n_t)$, $g(m_{1,t})$, and $g(m_{2,t})$ at their steady state values in (12). The steady state equilibrium conditions will force the eigenvalues $-(\delta + \theta - g(m_2))$ and $-(\delta - g(m_1))$ to be negative. And this will be consistent with a large gap between $g(m_2)$ and $g(m_1)$ if the rate θ 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

[10] What follows is a minimal description of this economy. The full specification appears in Section 6.

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 $g(m_2) - \delta_k > 0$ 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

$$\zeta = \min\left\{\frac{\delta_f}{[g(m_1) - \delta_k]^+}, \frac{\delta_f + \theta}{g(m_2) - \delta_k}\right\}. \tag{13}$$

The calculation is shown in Appendix A. US data imply that $\zeta > 1$ is close to 1. One possibility is that $g(m_1) - \delta_k > 0$ and that $\zeta = \delta_f/(g(m_1) - \delta_k)$ is close to 1. This means that $\delta_f + \delta_k - g(m_1) = \delta - g(m_1) > 0$ is close to zero. The alternative possibility is that $\zeta = (\delta_f + \theta)/(g(m_2) - \delta_k)$ is close to 1. This implies that $\delta_f + \delta_k + \theta - g(m_2) = \delta + \theta - g(m_2)$ 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 fact that large firms tend to have 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 is 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.

4. The Steady State and the Zipf Asymptote

Steady states are defined by $\mathrm{D}K_t = 0$ in (6) and $\mathrm{D}q_t = 0$ in (7), together with the static equilibrium conditions (8) and the capital-labor ratio implied by (4). 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 $A(k) = \alpha \in (0, 1)$ identically. This Cobb-Douglas economy implies a capital-labor ratio $k(\alpha)$ that can be computed from (4), resulting in a map $\alpha \mapsto k(\alpha)$. The steady state for an economy with a non-constant $A(\cdot)$ is simply a fixed point of the map $\alpha \mapsto A(k(\alpha))$. Assumption 2 is enough to guarantee a unique fixed point.

4.1 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

$$v = \frac{\mathrm{D}g(m)}{\rho + \delta - [g(m) - \mathrm{D}g(m)m]} \frac{1 - \alpha}{f(n[m])/(\delta - g(m))}, \tag{14}$$

$$D(v) = m \times \frac{f(n[m])}{\delta - g(m)} + n[m]. \tag{15}$$

Taking into account the steady state conditions $\mathrm{D}K_t = 0$ and $\mathrm{D}q_t = 0$, (14) follows from $v = qK\mathrm{D}g(m)/K$ and (15) 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 (14) is a strictly decreasing function of $m \in (0, m_\infty)$. It follows immediately that (14) defines v as a strictly decreasing function of $m \in (0, m_\infty)$. It is clear from (15) that $D(v)$ itself is strictly increasing in $m \in (0, m_\infty)$, and so (14)-(15) traces out a strictly decreasing demand curve for managerial services.

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

$$\alpha = (\beta v + w)L(v, w), \tag{16}$$

$$S(v) = M(v, w) - \beta L(v, w). \tag{17}$$

This supply curve is just a version of (5) with $A(k) = \alpha$ identically. As was shown in Lemma 3, it is strictly increasing.

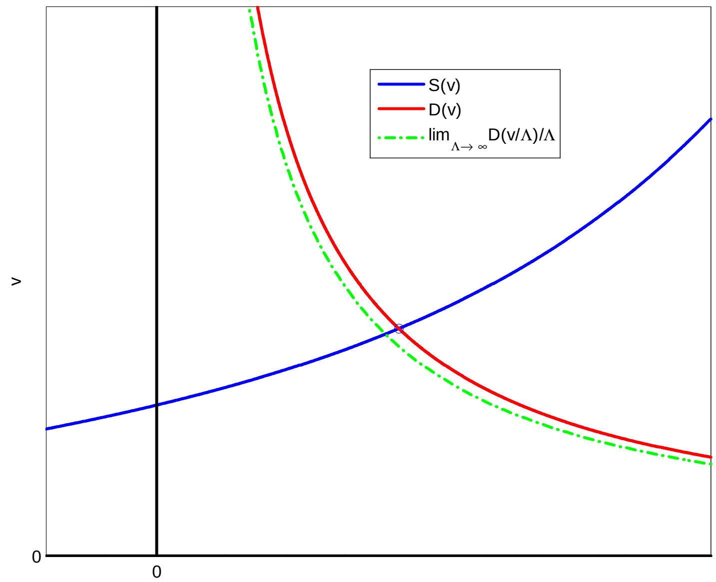

FIGURE 1 Steady State Demand and Supply for Managerial Services

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 (14)-(17), 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 (14) goes to ∞ as $m \downarrow 0$. The demand (15) 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 (14) 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 (14) is bounded away from zero. So the right-hand side of (14) 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 $v \in (0, \infty)$.■

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 (16) that an increase in α, holding fixed v, raises w. By (17), 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 (15) makes $D(v)$ an increasing function of m, and so a reduction in $D(v)$ has to go together with a reduction in m. This lowers the steady state capital stock $K = f(n[m])/(\delta - g(m))$. It turns out that $L(v, w)$ is increasing in α. A sketch of the proof is in Appendix B. So an increase in α lowers the capital-labor ratio, proving the following lemma.

Lemma 4 *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 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 (14)-(17) 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 (14)-(17) 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

$$\alpha' = A(k), \quad k = \frac{f(n[m])/(\delta - g(m))}{L(v, w)}. \tag{18}$$

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 $g(m) \uparrow \delta$.

4.3 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 a 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

$$M(v,w) = \Lambda M_1(\Lambda v, \Lambda w), \quad L(v,w) = \Lambda L_1(\Lambda v, \Lambda w), \tag{19}$$

where M_1 and L_1 are baseline supply curves that satisfy Assumption 1, and Λ is a positive parameter. One interpretation is that 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 then 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.

An alternative interpretation is simply that Λ measures the size of the population of households. Per-capita consumption is then C/Λ, and so $(\Lambda v, \Lambda w) = (\widetilde{v}, \widetilde{w})/(C/\Lambda)$ are the marginal utility weighted prices that matter for household choices. In either interpretation, the maintained assumption is that the fixed factor implicit in $f(\cdot)$ is not scaled by Λ.

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 (19). 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 (14)-(17) and let S_1 be the residual supply curve of managerial services implied by M_1 and L_1 and (16)-(17). 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 (14) 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, (14) implies

$$\lim_{\Lambda\to\infty} \Lambda[\delta - g(m)] = \left(\frac{(1-\alpha)\mathrm{D}g(m_\infty)}{\rho + \mathrm{D}g(m_\infty)m_\infty]}\right)^{-1} f(n[m_\infty])u. \tag{20}$$

The right-hand side of this equation is well defined and in $(0,\infty)$. From (15), the limiting scaled demand for managerial services is then

$$D_\infty(u) = \lim_{\Lambda\to\infty} D(u/\Lambda)/\Lambda = \frac{\mathrm{D}g(m_\infty)m_\infty}{\rho + \mathrm{D}g(m_\infty)m_\infty}\frac{1-\alpha}{u}. \tag{21}$$

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.[11] 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 $m \approx 0.22m_\infty$.

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 with $\Lambda\beta$ 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, unless F is approximately Cobb-Douglas for large capital-labor ratios). This would upset the argument used in Proposition 3 to prove that the derived demand for managerial services is an approximate hyperbola. Maintain, therefore, the factor supplies described in (19).

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 (19). Then the conclusions of Proposition 3 apply.*

[11] 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]).

Proof The proof 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 are given in the online 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 (19) 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 toward the process of replicating existing capital.

4.4 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.

Simplify by taking $\beta = 0$ and assuming that the large-Λ factor supplies implied by

(19) are determined by supply curves $L_1(v,w) = \mathcal{L}(w)$ and $M_1(v,w) = \mathcal{M}(v)$.[12] 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 (20)-(21). 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 $k = K/\mathcal{L}(w)$ 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 $-\mathrm{D}A(k)k/A(k) \leq 1 - A(k)$.[13] 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.

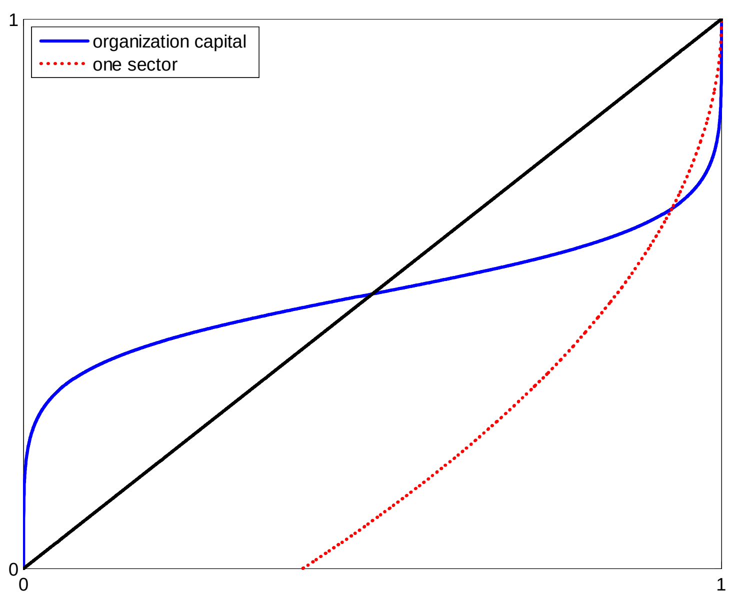

FIGURE 2 Existence When $A(k)$ Is Decreasing

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

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

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 $\rho + \delta = \mathrm{D}_1 F(k, 1)$. 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].

A Note on Prior Work The interpretation of Zipf's law implied by Propositions 3 and 4 is, in essence, the one given in Luttmer [2011]. But there one type of labor is used to produce consumption and replicate capital, while another type of labor is used to create capital from scratch. Here the human factors of production that are used to replicate capital and create capital from scratch are perfect substitutes and distinct from the labor that is also needed to produce consumption. It is clear that one can generalize and take all these human inputs to be imperfect substitutes in production.

5. Aggregate Dynamics

The pieces are now in place to examine aggregate convergence rates when replication and entry rates can respond to prices.

5.1 First Examples

Suppose $\beta = 0$ and that the supply of managerial services is completely inelastic, equal to a constant M. As long as the price of capital is positive, the net flow of new capital $\mathrm{D}K_t$ is then simply the maximum of $(g(m) - \delta)K_t + f(n)$ subject to the resource constraint $mK_t + n \leq M$. Managerial services are good for nothing else and are simply used to produce as much capital as possible. The first-order and envelope conditions immediately imply that

$$
\begin{aligned}
-\frac{\partial \mathrm{D}K_t}{\partial K_t} &= -\frac{\partial}{\partial K_t} \max_{m,n} \{(g(m)-\delta)K_t + f(n) : mK_t + n \le M\} \\
&= \delta - g(m_t) + \mathrm{D}g(m_t)m_t. \qquad (22)
\end{aligned}
$$

In the Zipf limit, $\delta = g(m_\infty)$, and so this reduces to $\mathrm{D}g(m_\infty)m_\infty = \delta\mathcal{S}_g$. Because a low capital stock implies more inelastically supplied managerial services per unit of capital, the capital stock converges to the steady state, even in the Zipf limit. The rate at which this happens is simply $\delta\mathcal{S}_g$. A 10% depreciation rate would imply half-lives of at least 7 years, and half-lives will be longer the lower is the factor share of managerial services.

This simple logic extends to scenarios with $\beta > 0$ when the supply of labor is also inelastic, at some $L < M/\beta$. In that case, the resource constraint is simply $mK_t + n \le M - \beta L$, and (22) again applies. Alternatively, if the supply of labor can respond to wages but the technology for producing consumption is determined by the Leontief production function $F(K, L) = \min\{K, L\}$, then the resource constraint becomes $(\beta + m)K_t + n \le M$. The first-order and envelope conditions then imply $-\partial \mathrm{D}K_t/\partial K_t = (\delta - g(m_t)) + \mathrm{D}g(m_t)(\beta + m_t)$. The resulting speed is therefore $\mathrm{D}g(m_\infty)(\beta + m_\infty) = (1 + \beta/m_\infty)\delta\mathcal{S}_g$ in the Zipf limit. A capital stock below the steady state releases managerial services that are normally used to produce consumption, and this will speed up convergence. Clearly, $\delta\mathcal{S}_g$ and $\beta/m_\infty = (\beta L/M)/(1 - \beta L/M)$ are critical parameters.[14]

How do these speeds change when $F(K, L)$ is Cobb-Douglas, or a CES technology with elasticity of substitution $\varepsilon \in (0, 1)$? And how do they compare to those of the Cass-Koopmans economy? To investigate the general case, it will be useful to restate the equilibrium conditions using the Hamiltonian for this economy.

5.2 The Hamiltonian

For positive K and q, define

$$
H(K, q) = \max_{(L,M,X)\in\mathcal{T}(K)} \{\ln(F(K, L)) - V(L, M) + qX\},
$$

where $\mathcal{T}(K)$ describes the technological constraints on accumulating capital,

$$
\mathcal{T}(K) = \left\{(L, M, X) \in \mathbb{R}^2_+ \times \mathbb{R} : X \le (g(m) - \delta)K + f(n), \quad mK + n + \beta L \le M\right\}.
$$

[14] If $F(K, L) = \min\{K/A, L/B\}$, then the span of control parameter that matters is $(\beta/m_\infty)B/A$.

It is not difficult to verify that the differential equation for (K_t, q_t) implied by (4)-(8) is the same as

$$\mathrm{D}K_t = \mathrm{D}_2 H(K_t, q_t), \quad \mathrm{D}q_t = \rho q_t - \mathrm{D}_1 H(K_t, q_t). \tag{23}$$

This Hamiltonian formulation of the equilibrium conditions follows Cass and Shell [1976].

Observe that $H(K_t, q_t)$ is concave in K_t and convex in q_t because $\ln(F(K, L))$ is concave and the graph of $K \mapsto \mathcal{T}(K)$ is convex. So (23) implies that $\partial \mathrm{D}K_t/\partial q_t$ and $\partial \mathrm{D}q_t/\partial K_t$ are both positive. Assumptions 1-3 ensure that $H(K_t, q_t)$ is sufficiently smooth to ensure that $\mathrm{D}_{12}H(K_t, q_t) = \mathrm{D}_{21}H(K_t, q_t)$. It then follows from (23) that $\partial \mathrm{D}K_t/\partial K_t$ and $\partial \mathrm{D}q_t/\partial q_t$ add up to ρ. Differentiating $(g(m_t) - \delta)K_t + f(n_t)$ with respect to K_t gives

$$\frac{\partial \mathrm{D}K_t}{\partial K_t} = g(m_t) - \delta + \mathrm{D}g(m_t) m_t \left(\frac{K_t}{m_t} \frac{\partial m_t}{\partial K_t} + \frac{n_t}{m_t K_t} \frac{K_t}{n_t} \frac{\partial n_t}{\partial K_t} \right),$$

which generalizes (22). The last term on the right-hand side is negative because of Lemma 3 and the elasticities (9)-(10) implied by the static equilibrium conditions. So $\partial \mathrm{D}K_t/\partial K_t < 0$ when $\delta > g(m_t)$. The following lemma collects these results.

Lemma 5 *Suppose Assumptions 1-3 hold and consider the function* $(K_t, q_t) \mapsto (\mathrm{D}K_t, \mathrm{D}q_t)$ *defined by (23). Then* $\partial \mathrm{D}K_t/\partial K_t + \partial \mathrm{D}q_t/\partial q_t = \rho$ *and* $\partial \mathrm{D}K_t/\partial q_t > 0$ *and* $\partial \mathrm{D}q_t/\partial K_t > 0$. *Furthermore,* $\partial \mathrm{D}K_t/\partial K_t < 0$ *whenever* $\delta > g(m_t)$.

Let $-\mathcal{D}$ be the determinant of $\partial[\mathrm{D}K_t, \mathrm{D}q_t]/\partial[K_t, q_t]$. Using the fact that $\partial \mathrm{D}q_t/\partial q_t = \rho - \partial \mathrm{D}K_t/\partial K_t$ gives

$$\mathcal{D} = \frac{\partial \mathrm{D}K_t}{\partial q_t} \frac{\partial \mathrm{D}q_t}{\partial K_t} - \frac{\partial \mathrm{D}K_t}{\partial K_t} \left(\rho - \frac{\partial \mathrm{D}K_t}{\partial K_t} \right). \tag{24}$$

The eigenvalues of $\partial[\mathrm{D}K_t, \mathrm{D}q_t]/\partial[K_t, q_t]$ are given by $-(\rho/2) \pm \sqrt{(\rho/2)^2 + \mathcal{D}}$. Since $0 = \mathrm{D}K_t/K_t \geq g(m_t) - \delta$ in the steady state, an immediate consequence of Lemma 5 is that $\mathcal{D}$ is positive when evaluated at the steady state. The eigenvalues of $\partial[\mathrm{D}K_t, \mathrm{D}q_t]/\partial[K_t, q_t]$ are then real and of opposite signs. That is, the steady state is a saddle point. In a neighborhood of the steady state, there will be a stable manifold with a slope implied by the eigenvector associated with the negative eigenvalue of $\partial[\mathrm{D}K_t, \mathrm{D}q_t]/\partial[K_t, q_t]$. This slope is

$$\frac{\partial q_t}{\partial K_t} = -\frac{\frac{\partial \mathrm{D}q_t}{\partial K_t}}{\frac{\rho}{2} - \frac{\partial \mathrm{D}K_t}{\partial K_t} + \sqrt{\left(\frac{\rho}{2}\right)^2 + \mathcal{D}}}, \tag{25}$$

which is negative by Lemma 5. The rate at which (K_t, q_t) converges to the steady state along the stable manifold is

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

This is strictly increasing in $\mathcal{D}$ and equals $\mathcal{D}$ to a first-order approximation when $\mathcal{D}$ is small. We need to study how $\mathcal{D}$ depends on the underlying parameters of this economy.

5.3 The Speed of Convergence

An explicit computation of $\partial[\mathrm{D}K_t, \mathrm{D}q_t]/\partial[K_t, q_t]$ gives

$$\begin{aligned}
\begin{bmatrix} \frac{\partial \mathrm{D}K_t}{\partial K_t} & \frac{q_t}{K_t}\frac{\partial \mathrm{D}K_t}{\partial q_t} \\ \frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t} & \frac{\partial \mathrm{D}q_t}{\partial q_t} \end{bmatrix} &= \begin{bmatrix} -(\delta - g(m)) & 0 \\ \rho + \delta - g(m) & \rho + \delta - g(m) \end{bmatrix} \\
&+ \mathcal{S}_g g(m) \begin{bmatrix} \frac{K}{m}\frac{\partial m}{\partial K} + \frac{n}{mK}\frac{K}{n}\frac{\partial n}{\partial K} & \frac{q}{m}\frac{\partial m}{\partial q} + \frac{n}{mK}\frac{q}{n}\frac{\partial n}{\partial q} \\ 1 - \mathcal{C}_g \times \frac{K}{m}\frac{\partial m}{\partial K} & 1 - \mathcal{C}_g \times \frac{q}{m}\frac{\partial m}{\partial q} \end{bmatrix} \\
&+ (\rho + \mathcal{S}_g\delta + (1 - \mathcal{S}_g)(\delta - g(m))) \times \frac{A(k)\mathcal{E}_A}{1 - A(k)} \begin{bmatrix} 0 & 0 \\ \frac{K}{k}\frac{\partial k}{\partial K} & \frac{q}{k}\frac{\partial k}{\partial q} \end{bmatrix}
\end{aligned} \tag{27}$$

The elasticities on the right-hand side of (27) are the elasticities calculated in (9)-(11). Note well that $\delta - g(m)$ is positive in the steady state. Combining (27) with the expression (24) for the approximate speed leads to some basic observations of the speed of convergence of this economy.

Observation 1 *The off-diagonal $\partial \mathrm{D}q_t/\partial K_t > 0$ is increasing in the elasticity $\mathcal{E}_A \geq 0$. Holding fixed the static elasticities of (m, n, k) with respect to (K, q), the closer the production function in the consumption sector is to Cobb-Douglas, the slower the economy converges to its steady state.*

Observation 2 *Both $-\partial \mathrm{D}K_t/\partial K_t$ and the off-diagonal elements of (27) are increasing in $\delta - g(m)$ and $\mathcal{S}_g$. Other things equal, lowering $\delta - g(m) > 0$ and the factor share $\mathcal{S}_g \in (0, 1)$ reduces the speed of convergence.*

The simple intuition for the first of these observations is that $\mathcal{E}_A > 0$ and $\partial k_t/\partial K_t > 0$ means that profits per unit of capital are high when K_t is low, thus increasing the incentives to accumulate organization capital and raising the convergence rate of the economy. The second observation is anticipated by (22). It implies that the slow convergence property described in the Introduction and Section 3 survives if $\mathcal{S}_g$ and $\mathcal{E}_A$ are small. But for any given $\mathcal{S}_g > 0$ and $\mathcal{E}_A > 0$, the speed of convergence does not literally go to zero as $g(m) \uparrow \delta$.

5.3.1 The Zipf Limit

The terms $\delta - g(m)$ disappear from (27) in the $g(m) \uparrow \delta$ limit, and mK becomes large relative to n. So the coefficient $n/(mK)$ in (27) 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 (9)-(10) converges to

$$\begin{bmatrix} \frac{K}{m}\frac{\partial m}{\partial K} & \frac{q}{m}\frac{\partial m}{\partial q} \\ \frac{K}{n}\frac{\partial n}{\partial K} & \frac{q}{n}\frac{\partial n}{\partial q} \end{bmatrix} = \begin{bmatrix} 1/\mathcal{C}_g \\ 1/\mathcal{C}_f \end{bmatrix} \frac{\begin{bmatrix} \mathcal{E}_{S,K} - 1 & \mathcal{E}_{S,v} \end{bmatrix}}{\mathcal{E}_{S,v} + 1/\mathcal{C}_g}. \tag{28}$$

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 (27) because the contribution of new capital created from scratch becomes negligible.[15] 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 raise m a bit, but the resulting steep reduction in the marginal product $\mathrm{D}g(m)$ implied by a high $\mathcal{C}_g$ mostly reduces the supply of managerial services.

Using $g(m) \uparrow \delta$ together with (28) to simplify (27) yields

$$\begin{aligned} \begin{bmatrix} \frac{\partial \mathrm{D}K_t}{\partial K_t} & \frac{q_t}{K_t}\frac{\partial \mathrm{D}K_t}{\partial q_t} \\ \frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t} & \frac{\partial \mathrm{D}q_t}{\partial q_t} \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ \rho & \rho \end{bmatrix} + \frac{\mathcal{S}_g \delta}{1 + \mathcal{C}_g \mathcal{E}_{S,v}} \begin{bmatrix} -(1 - \mathcal{E}_{S,K}) & \mathcal{E}_{S,v} \\ 1 + \mathcal{C}_g(\mathcal{E}_{S,v} + 1 - \mathcal{E}_{S,K}) & 1 \end{bmatrix} \\ &\quad + (\rho + \mathcal{S}_g \delta) \times \frac{A(k)\mathcal{E}_A}{1 - A(k)} \begin{bmatrix} 0 & 0 \\ \frac{K}{k}\frac{\partial k}{\partial K} & \frac{q}{k}\frac{\partial k}{\partial q} \end{bmatrix}. \end{aligned} \tag{29}$$

[15] With entry out of the picture, the economy starts to look like the two-sector economy of Uzawa [1961]. The difference is that here, capital can be used to produce consumption and new capital at the same time.

The implied approximate speed of convergence (24) is then equal to

$$\mathcal{D} = \frac{(\rho + \delta\mathcal{S}_g)\,\delta\mathcal{S}_g}{\frac{1}{1+\mathcal{E}_{S,v}} + \frac{\mathcal{E}_{S,v}}{1+\mathcal{E}_{S,v}} \times \mathcal{C}_g} \times \mathcal{F}, \tag{30}$$

where

$$\mathcal{F} = \frac{\mathcal{E}_{S,v}}{1+\mathcal{E}_{S,v}} \left(1 + \frac{A(k)\mathcal{E}_A}{1-A(k)} \frac{K}{k} \frac{\partial k}{\partial K}\right) + \frac{1-\mathcal{E}_{S,K}}{1+\mathcal{E}_{S,v}} \left(1 + \frac{\delta\mathcal{S}_g}{\rho+\delta\mathcal{S}_g} \frac{-\mathcal{E}_{S,K}}{1+\mathcal{C}_g\mathcal{E}_{S,v}}\right). \tag{31}$$

The residual managerial supply elasticities $\mathcal{E}_{S,K} \leq 0$ and $\mathcal{E}_{S,v} > 0$ were defined in Lemma 3, and the partial elasticity $(K/k)\partial k/\partial K$ implied by the static equilibrium conditions was reported in (11). Note that an approximate speed $(\rho + \delta\mathcal{S}_g)\,\delta\mathcal{S}_g$ combined with (26) results in the exact speed $\delta\mathcal{S}_g$ obtained in Section 5.1 (the Zipf version of (22)). Recall that Assumption 2 ensures that $-\mathcal{E}_{S,K} \geq 0$ and $\mathcal{E}_A \geq 0$, and both elasticities are zero if the technology for producing consumption is Cobb-Douglas. So the first factor on the right-hand side of (30) determines the Cobb-Douglas speed, and $\mathcal{F}$ is an adjustment factor for deviations from a Cobb-Douglas technology. Also, (11) implies that $(K/k)\partial k/\partial K$ is positive. So the assumption that the consumption-sector labor share $A(k)$ is non-decreasing implies $\mathcal{F} \geq 1$. The economy converges more quickly than the Cobb-Douglas economy if $A(\cdot)$ is strictly increasing.[16]

In the Zipf limit, the first-order condition for m can be written as $vmK = qK\delta\mathcal{S}_g$, and the steady state condition for q reduces to $qK = (1 - A(k))/(\rho + \delta\mathcal{S}_g)$. Using $A(k) = (\beta v + w)L(v,w)$ and $M(v,w) = \beta L(v,w) + mK$, it follows that

$$\frac{\delta\mathcal{S}_g}{\rho+\delta\mathcal{S}_g} \frac{\frac{\beta L}{M}}{1-\frac{\beta L}{M}} = \frac{\beta v}{\beta v + w} \frac{A(k)}{1-A(k)}. \tag{32}$$

This equation can be used to express the deviation $\mathcal{D}/\left(\rho + \delta\mathcal{S}_g\right)\delta\mathcal{S}_g$ from the Zipf version of (22) entirely in terms of steady state factor shares and elasticities, independently of the discount rate ρ. The high-level elasticity $(q/k)(\partial k/\partial q)$ does not appear explicitly in (30)-(31) because of a judicious use of $\partial \mathrm{D}q_t/\partial q_t = \rho - \partial \mathrm{D}K_t/\partial K_t$. But the elasticity $(K/k)\partial k/\partial K$ remains, and it is a rather unwieldy function of more primitive elasticities. The Cobb-Douglas case and the case of separable factor supplies offer further simplification.

[16] More precisely, this comparison applies holding fixed the elasticity $\mathcal{E}_{S,v}$, which itself depends on $\mathcal{E}_A$ and other parameters, by Lemma 2 and (5).

5.3.2 More on the Cobb-Douglas Case

As noted, the first factor on the right-hand side of (30) determines the approximate speed for an economy in which the technology for the consumption sector is Cobb-Douglas. A replication technology with more curvature implies a slower speed. In the special case of $\mathcal{C}_g = 1$, the Cobb-Douglas approximate speed implies the same exact speed $\delta\mathcal{S}_g$ as in the simple economy discussed in Section 5.1 (the Zipf version of (22)), independently of the residual supply elasticity $\mathcal{E}_{S,v}$. The economy converges more quickly than $\delta\mathcal{S}_g$ when $\mathcal{C}_g \in (0,1)$, as is the case, for example, if $g(m)$ is also Cobb-Douglas (so that $\mathcal{C}_g = 1 - \mathcal{S}_g$.) The right-hand side of (30) is then an increasing function of $\mathcal{E}_{S,v}$. The speed of convergence declines toward $\delta\mathcal{S}_g$ as $\mathcal{E}_{S,v} \downarrow 0$. The convergence rate then goes to zero as the managerial factor share $\mathcal{S}_g$ becomes small.

A different slow convergence scenario arises when $\mathcal{C}_g$ is large. The maintained assumption that $\mathcal{E}_{S,v} > 0$ ensures that the supply of managerial services will respond to a low factor price 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 approaches 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 quite 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. When managers have other options, the extent to which m_t can rise when K_t falls below the steady state is limited.

5.3.3 Separable Factor Supplies

To extract some further implications from the speed formula (30)-(31), suppose $M(v, w) = M(v)$ and $L(v, w) = L(w)$. That is, the factor supplies are separable. One can imagine that there are two types of household members: worker types who value leisure and have an ability to supply labor, and manager types who also value leisure but have an ability to supply managerial services. This simplifies the expressions for the elasticities $\mathcal{E}_{S,K}$, $\mathcal{E}_{S,v}$, and $(K/k)\partial k/\partial K$ that appear in (30)-(31). A somewhat elaborate calculation gives an approximate speed determined by

$$\mathcal{D} = \frac{(\rho+\delta\mathcal{S}_g)\delta\mathcal{S}_g}{\frac{1}{1+\mathcal{E}_{S,v}} + \frac{\mathcal{E}_{S,v}}{1+\mathcal{E}_{S,v}} \times \mathcal{C}_g} \times \left(1 + \mathcal{E}_A \times \frac{\frac{A(k)}{1-A(k)}\left(1+\frac{w}{\beta v+w}\frac{1}{\mathcal{E}_L}\right)\mathcal{E}_M + \frac{\beta L}{M} + \frac{A(k)}{1-A(k)}\frac{\beta v}{\beta v+w}}{\left(1+\mathcal{E}_A+\frac{w}{\beta v+w}\frac{1}{\mathcal{E}_L}\right)\left(1-\frac{\beta L}{M}+\mathcal{E}_M\right)+\frac{\beta L}{M}\frac{\beta v}{\beta v+w}}\right) \tag{33}$$

with a residual supply elasticity of managerial services equal to

$$\mathcal{E}_{S,v} = \frac{1}{1-\frac{\beta L}{M}}\left(\mathcal{E}_M + \frac{\frac{\beta L}{M}\frac{\beta v}{\beta v+w}}{1+\mathcal{E}_A+\frac{w}{\beta v+w}\frac{1}{\mathcal{E}_L}}\right). \tag{34}$$

This allows one to compute the speed of this economy solely on the basis of the share parameters $A(k)$, $\beta v/(\beta v+w)$, and $\beta L/M$, the curvature $\mathcal{C}_g$ of the replication technology, and the primitive elasticities $\mathcal{E}_L$, $\mathcal{E}_M$, and $\mathcal{E}_A$. The limits $\mathcal{E}_L \to \infty$ and $\mathcal{E}_M \downarrow 0$ simplify (33)-(34) and leave $\mathcal{E}_A$ as the only elasticity to be estimated for a back-of-the-envelope calculation. The first factor on the right-hand side of (33) is the Cobb-Douglas approximate speed for a given $\mathcal{E}_{S,v}$. It is now the only component of (33) that depends on $\mathcal{C}_g$. As expected, $\mathcal{E}_A > 0$ implies speeds that are slower than those in a Cobb-Douglas economy with the same $\mathcal{E}_{S,v}$. Clearly, $\mathcal{E}_L \downarrow 0$ and $\mathcal{E}_M \downarrow 0$ imply $\mathcal{D} \to (\rho+\delta\mathcal{S}_g)\delta\mathcal{S}_g$, resulting in the exact speed $\delta\mathcal{S}_g$ described in Section 5.1. Taking $\mathcal{E}_A \to \infty$ gives $\mathcal{E}_{S,v} \to \mathcal{E}_M/(1-\beta L/M)$ in (34) and then (33) yields

$$\mathcal{D} \to \frac{(\rho+\delta\mathcal{S}_g)\delta\mathcal{S}_g}{1-\frac{\beta L}{M}+\mathcal{C}_g\mathcal{E}_M}\left(\frac{1+\mathcal{E}_M}{1-A(k)} + \frac{A(k)}{1-A(k)}\frac{w}{\beta v+w}\frac{\mathcal{E}_M-\mathcal{E}_L}{\mathcal{E}_L}\right). \tag{35}$$

The factor share $A(k)$ for this Leontief limit is determined by (32). The Leontief speed (35) is decreasing in the labor supply elasticity $\mathcal{E}_L$. Intuitively, a destruction of capital in the Leontief economy forces a reduction in labor supply, and this requires only a small change in marginal utility weighted wages when $\mathcal{E}_L$ is substantial. This limits the extent to which marginal utility weighted Leontief profits can rise to help speed up a recovery. One can verify that $\mathcal{E}_M = 0$ produces the exact speed $(1+\beta/m_\infty)\delta\mathcal{S}_g$ obtained in Section 5.1. How the speed of convergence varies with $\mathcal{E}_M > 0$ depends on $\mathcal{C}_g$. For small values of this curvature parameter, a more elastic supply of managerial services increases the speed. But high curvature produces a negative dependence on $\mathcal{E}_{S,v} \propto \mathcal{E}_M$, as it does in the Cobb-Douglas case.

5.4 Relation to Cass-Koopmans

It is interesting to see how these results relate to a Cass-Koopmans economy with logarithmic utility and an aggregate labor supply $M(w,w) + L(w,w)$. In such an economy, $F(K,L) = Kg(L/K)$ and the capital stock evolves according to $\mathrm{D}K_t = (g(m_t) - \delta)K_t - C_t$. The economy produces one type of output that can be thought of as new capital, and households can consume out of the "inventory" of productive 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 $q_t = 1/C_t$ 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 (7), and so the asset pricing equation for q_t is simply $\mathrm{D}q_t = (\rho + \delta - (g(m_t) - \mathrm{D}g(m_t)m_t))\, q_t$. This implies $\rho + \delta = (1 - \mathcal{S}_g)g(m)$ in a steady state.

5.4.1 Not $g(m)/\delta < 1$ but $g(m)/\delta > 1$

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

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

where $g(m) = (\rho + \delta)/(1 - \mathcal{S}_g) > \delta$. Gross output is $g(m)K$, 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 $\mathcal{C}_g/(1 - \mathcal{S}_g)$ is the reciprocal of the elasticity of substitution between capital and labor. It disappears from (36) if the technology is Cobb-Douglas. If $\mathcal{E} = \infty$, as in Rogerson [1988], then the Cobb-Douglas version of (36) reduces to simply $(\rho + \delta)\mathcal{S}_g/(1 - \mathcal{S}_g) \approx 0.33$, which implies a half-life of approximately 2.2 years. If $\mathcal{E}$ happens to equal $\mathcal{E}_{S,v}$ and the same Cobb-Douglas technology is used in both (30) and (36), then the last factor that appears in the discriminant of (36) corresponds to $\mathcal{D}$ with $\mathcal{F} = 1$, given in (30). But at $g(m)/\delta \approx 5$ instead of slightly below 1, (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 yields a convergence rate of 0.10 in the economy with organization capital, and a bit more than 0.24 in the Cass-Koopmans economy. The implied half-lives are 6.8 years and 2.9 years, respectively.

5.4.2 One versus Two Types of Curvature

A striking difference between (30) 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 generates long-run growth[17]—that is, no convergence at all—if $\rho+\delta$ is low enough, though with implications for the labor share that appear, so far, to be completely counterfactual.

The economy with organization capital has two curvature parameters that matter for the speed of convergence. 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 relatively fast convergence—the factor $\mathcal{F}$ in (30)-(31) 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 (30) 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.

[17]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 $\mathcal{S}_k + \mathcal{S}_l = 1$. So given share parameters, high curvature of $F(k,1)$ is the same as high curvature of $F(1,l)$.

6. Quantitative Results for a Two-Type Economy

Given the evidence on persistent heterogeneity in firm growth rates, it is important to investigate how such heterogeneity affects the speed calculations obtained so far. And it is important to learn whether the equilibrium trajectories of some version of this economy resemble what happens in recessions followed by slow recoveries. The two-type economy sketched in Section 3.1 can be used for this purpose.

Consider again an economy with two types of capital, indexed by $j \in \{1, 2\}$, that vary in quality. The technology for producing consumption is that of a vintage capital economy: the consumption produced by $K_{j,t}$ units of type-j capital and $L_{j,t}$ units of team services is $z_j F(K_{j,t}, L_{j,t})$, where $z_2 > z_1 > 0$. The flow of new type-j capital produced from scratch is $\phi_j f(n_t)$, where ϕ_j is a probability in $(0, 1)$. Entrants cannot guarantee up front what the quality of the new capital they produce will be. As before, capital can be replicated at the rate $g(m_{j,t})$. But now assume that the marginal product of $g(\cdot)$ is finite at zero. This introduces the possibility that low-quality capital is not replicated at all, while at the same time, high-quality capital is replicated at rates near a finite speed limit $g(\infty)$. Firms exit randomly, and all their capital is destroyed, at the rate δ_f. Capital also depreciates physically at the rate $\delta_k = \delta - \delta_f$. In addition, firms with only type-2 capital become firms with only type-1 capital randomly at the rate θ.

The stationary distribution of firm size was given in Section 3.1. But an explicit characterization of the speed of convergence is not as easy as in the one-type economy. The Zipf limit again offers a significant simplification, and a complete characterization of the dynamic properties of this economy is possible when F is Leontief. The dynamic properties of economies with CES consumption technologies with low but positive elasticities of substitution can then be understood by continuity. In particular, the Leontief benchmark turns out to give a reasonable estimate of the convergence speed of an economy in which the elasticity of substitution of F is greater than zero but not too large.

6.1 The Zipf Steady State

Consider the Zipf limiting economy obtained by letting factor supplies become large. Suppose the equilibrium path is such that $v_t = q_{2,t}\mathrm{D}g(m_{2,t}) > q_{1,t}\mathrm{D}g(0)$, so that only type-1 capital is replicated. The capital stocks then evolve according to $\mathrm{D}K_{1,t} = -\delta K_{1,t} + \theta K_{2,t}$ and $\mathrm{D}K_{2,t} = -(\delta + \theta)K_{2,t} + g(m_{2,t})K_{2,t}$. A steady state then requires that $\delta K_1 =$

θK_2 and $m_2 = m_\infty$, with m_∞ now defined to be the solution to $\delta + \theta = g(m_\infty)$. Given capital stocks K_1 and K_2, and factor supplies L and $M > \beta L$, the aggregate supply of consumption goods is given by

$$C(K_1, K_2, L) = \max_{L_1, L_2} \{z_1 F(K_1, L_1) + z_2 F(K_2, L_2) : L_1 + L_2 \leq L\}. \qquad (37)$$

This production function summarizes all the information about z_1, z_2, and F needed for determining an equilibrium.[18] Assume F is sufficiently smooth. The first-order and envelope conditions imply that the price of a team of managers and workers will have to be $(\beta v + w)C(K_1, K_2, L) = \mathrm{D}_3 C(K_1, K_2, L)$. It is not difficult to verify that $\mathrm{D}_3 C(K_1, K_2, L)$ is increasing in both K_1 and K_2. Profits per unit of type-j capital are equal to $\mathrm{D}_j C(K_1, K_2, L)$ units of consumption, and these are both decreasing in (K_1, K_2) and increasing in L. Given that only type-2 capital is replicated, and taking into account that $\theta K_2 = \delta K_1$, steady state capital prices are determined by the present value conditions

$$\begin{aligned} q_1 K_1 &= \frac{1}{\rho + \delta} \frac{\mathrm{D}_1 C(K_1, K_2, L) K_1}{C(K_1, K_2, L)}, \\ q_2 K_2 &= \frac{1}{\rho + (\delta + \theta)\mathcal{S}_g} \left(\frac{\delta}{\rho + \delta} \frac{\mathrm{D}_1 C(K_1, K_2, L) K_1}{C(K_1, K_2, L)} + \frac{\mathrm{D}_2 C(K_1, K_2, L) K_2}{C(K_1, K_2, L)} \right), \end{aligned}$$

where $\mathcal{S}_g = \mathrm{D}g(m_\infty) m_\infty / g(m_\infty)$. Because type-1 capital is not being replicated, the effective discount rate applied to the factor share of type-1 capital is simply $\rho + \delta$. The first term on the right-hand side of the equation for $q_2 K_2$ accounts for the fact that type-2 capital depreciates into type-1 capital at the rate θ, as well as the fact that $\theta K_2 = \delta K_1$ in any steady state. Because of replication, the effective discount rate for type-2 profits is really $\rho + \delta + \theta - (g(m_\infty) - \mathrm{D}g(m_\infty) m_\infty)$, but the Zipf limit implies that the term $\delta + \theta - g(m_\infty)$ drops out.

Using the equation for $q_2 K_2$ to eliminate q_2 from the usual optimality condition for replication, $v = q_2 \mathrm{D}g(m_\infty)$, gives

$$\beta v L = \frac{(\delta + \theta)\mathcal{S}_g}{\rho + (\delta + \theta)\mathcal{S}_g} \left(\frac{\delta}{\rho + \delta} \frac{\mathrm{D}_1 C(K_1, K_2, L) K_1}{C(K_1, K_2, L)} + \frac{\mathrm{D}_2 C(K_1, K_2, L) K_2}{C(K_1, K_2, L)} \right) \frac{\beta L}{m_\infty K_2}. \qquad (38)$$

[18] If F is Cobb-Douglas, this will be a Cobb-Douglas function of labor and a quality-weighted aggregate of the two capital stocks. But this aggregation result is very special. In particular, it fails for all other CES production functions.

This is the factor share of managerial services in producing consumption. Note that the right-hand side of (38) is a function only of L/K_2, given the steady state condition $\delta K_1 = \theta K_2$. A proportional increase in both K_1 and K_2 (as required by $\delta K_1 = \theta K_2$) lowers both marginal products $\mathrm{D}_1 C(K_1, K_2, L)$ and $\mathrm{D}_2 C(K_1, K_2, L)$. The right-hand side of (38) is therefore increasing in L/K_2, taking into account $\delta K_1 = \theta K_2$. In other words, more labor unambiguously raises the managerial factor share in the consumption sector.[19]

The overall team cost share in consumption is $(\beta v + w)L$, and the first-order and envelope conditions for (37) imply

$$(\beta v + w)L = \frac{\mathrm{D}_3 C(K_1, K_2, L)L}{C(K_1, K_2, L)}. \tag{39}$$

This is also a function only of L/K_2, taking into account $\delta K_1 = \theta K_2$. In contrast to the factor share (38), the factor share (39) may rise or fall with L/K_2, even if the factor shares of F happen to be monotone in the capital-labor ratio. The remaining equilibrium conditions are the resource constraints

$$L = L(v, w), \quad M(v, w) = \beta L + m_\infty K_2, \quad \delta K_1 = \theta K_2. \tag{40}$$

The conditions (38)-(40) determine the steady state for this economy, provided that $v \geq q_1 \mathrm{D}g(0)$, so that $m_1 = 0$ is optimal. Using (38) and (39) to eliminate v and w from (40) gives two equilibrium conditions that can be used to determine L and K_2. Observe that one can verify whether (K_1, K_2, L) combined with (v, M) and (w, L) is a steady state knowing only factor shares, and without knowledge of elasticities of substitution or factor supply elasticities.

6.1.1 Separable Factor Supplies

Suppose now that the factor supplies are of the form $M(v)$ and $L(w)$. This assumption makes it particularly easy to infer factor prices from factor supplies. Combining (38)-(39)

[19] None of this makes any assumptions about F other than that it is concave and exhibits constant returns to scale.

and (39)-(40) then gives

$$\begin{aligned}\frac{(\delta+\theta)\mathcal{S}_g}{\rho+(\delta+\theta)\mathcal{S}_g}\left(\frac{\theta}{\rho+\delta}\frac{\mathrm{D}_1C(K_1,K_2,L)}{\mathrm{D}_3C(K_1,K_2,L)}+\frac{\mathrm{D}_2C(K_1,K_2,L)}{\mathrm{D}_3C(K_1,K_2,L)}\right)\frac{\beta}{m_\infty} & \\ +\frac{L^{-1}(L)C(K_1,K_2,L)}{\mathrm{D}_3C(K_1,K_2,L)} &= 1 \quad (41) \\ M\left(\frac{1}{\beta}\left(\frac{\mathrm{D}_3C(K_1,K_2,L)}{C(K_1,K_2,L)}-L^{-1}(L)\right)\right)-\beta L &= m_\infty K_2. \quad (42)\end{aligned}$$

So (41) equates the price of a team with its marginal product, taking into account the wages implied by the labor supply curve, and taking into account how the factor price of managerial services is determined by the present value of future profits. The two terms on the left-hand side of (41) represent $\beta v/(\beta v+w)$ and $w/(\beta v+w)$, respectively. Clearly, this is a curve that can be shifted by changes in the discount rate ρ, but not by changes in the supply curve of managerial services. Equation (42) uses the presumed equality of team price and marginal product to back out the factor price of managerial services and then imposes the market clearing condition for managerial services. This curve does not depend on ρ but will shift when factor supply curves change.

The left-hand side of (41) is strictly increasing in L, and that of (42) is strictly decreasing. If $L(0)=0$ and $L(\infty)\in(0,\infty)$, then both (41) and (42) will have unique solutions as long as the marginal products of $C(K_1,K_2,L)$ are in $(0,\infty)$. With $\theta K_2=\delta K_1$ understood, this then defines two functions $K_2\mapsto L$. Examples are shown in Figures 3 and 6.

6.1.2 The Leontief Benchmark

To simplify further, consider the Leontief technology $F(K,L)=\min\{K,L\}$. Given $z_2>z_1$ and $M>\beta L$, aggregate consumption will be

$$C(K_1,K_2,L)=z_1\min\{K_1+K_2,L\}+(z_2-z_1)\min\{K_2,L\}. \quad (43)$$

The high-quality type-2 capital is used first. Low-quality type-1 capital is used only if the supply of labor exceeds K_2. So $C(K_1,K_2,L)$ is piecewise linear in L, with kinks at $L=K_2$ and $L=K_1+K_2$. The efficiency conditions (38)-(39) have to be restated in terms of tangency conditions. The factor price of a team of managers and workers now has to satisfy

$$\begin{aligned}(\beta v + w)C &= z_2 && \text{if } L \in (0, K_2),\\ (\beta v + w)C &\in [z_1, z_2] && \text{if } L = K_2,\\ (\beta v + w)C &= z_1 && \text{if } L \in (K_2, K_2 + K_1),\\ (\beta v + w)C &\in [0, z_1] && \text{if } L = K_1 + K_2,\end{aligned} \tag{44}$$

where $C = C(K_1, K_2, L)$. The resulting profits per unit of type-j capital are $\max\{0, z_j - (\beta v + w)C\}$ units of consumption, and capital prices are then determined by

$$q_1 C = \frac{\max\{0, z_1 - (\beta v + w)C\}}{\rho + \delta}, \quad q_2 C = \frac{\max\{0, z_2 - (\beta v + w)C\} + \theta q_1 C}{\rho + (\delta + \theta)\mathcal{S}_g}, \tag{45}$$

again, with $C = C(K_1, K_2, L)$. The remaining equilibrium conditions are the first-order condition $v = q_2 \mathrm{D}g(m_\infty)$ and the resource constraints (40). This describes a candidate steady state, and it will be an actual steady state if $v \geq q_1 \mathrm{D}g(0)$.

Suppose now that managerial services are supplied inelastically at some level M and that the supply of labor is $L = L(w)$. Suppose $1 = wL(w)$ implies $M < (\beta + m_\infty)L(w)$. This ensures that managerial services are scarce, ruling out a steady state in which abundant managerial services can be used to ensure $L(w) < K_2$ at any feasible wage. The steady state can then be characterized as follows.

Proposition 5 *Consider the Zipf limit of a two-type Leontief economy in which low-quality capital is not replicated and managerial services are scarce. Define m_∞ to be the solution to $g(m_\infty) = \delta + \theta$ and assume that*

$$\frac{\beta}{m_\infty} \frac{z_2 - z_1}{z_1} < 1. \tag{46}$$

Then the steady state of this economy is determined by a tangency condition,

$$\mathcal{B} + L^{-1}(L)\left(L + \left(\frac{z_2 - z_1}{z_1}\right) K_2\right) \in \begin{cases} [1, \infty), & L = K_2, \\ \{1\}, & L \in (K_2, K_2 + K_1), \\ [0, 1], & L = K_2 + K_1, \end{cases} \tag{47}$$

where $\delta K_1 = \theta K_2$, and a resource constraint,

$$M = \beta L + m_\infty K_2. \tag{48}$$

The parameter

$$\mathcal{B} = \frac{(\delta+\theta)\mathcal{S}_g}{\rho+(\delta+\theta)\mathcal{S}_g}\frac{\beta}{m_\infty}\frac{z_2-z_1}{z_1} \tag{49}$$

is the cost share $\beta v/(\beta v{+}w)$ of managerial supervision in the scenario $L \in (K_2, K_2{+}K_1)$. In this scenario, a reduction in ρ that is not too large increases the capital stocks K_1 and K_2 but lowers L, $w = L^{-1}(L)$, and C. Steady state flow utility $\ln(C) - V(L, M)$ is maximized at the steady state associated with $\rho = 0$.

The tangency condition (47) is a version of (41), and the resource constraint (48) is the inelastic version of (42). Note that in the scenario $L \in (K_2, K_2 + K_1)$, $C = z_1 L + (z_2 - z_1)K$ and $M = \beta L + m_\infty K$ imply $C = z_1 L + (z_2 - z_1)(M - \beta L)/m_\infty$, which is increasing in L if and only if (46) holds. Condition (46) says that the steady state production possibility frontier for $(L(\infty) - L, C)$ has a segment where K_1 is actually used to produce consumption. If $\beta > 0$, this can only happen if the quality difference between the two types of capital is not too large. Condition (46) implies that the resource constraint $M = \beta L + m_\infty K_2$ has a steeper slope $\partial L/\partial K_2$ than the tangency condition (48) evaluated at any $L \in (K_2, K_2 + K_1)$. The comparative statics for ρ follow easily from this observation.

Figure 3 shows the $L = K_1 + K_2$ scenario for this Leontief economy. The curve contained in the cone $L/K_2 \in [1, 1{+}\theta/\delta]$ and labeled "present value" and "$\varepsilon = 0$" is (47). The downward-sloping line is (48). An expansion of the supply of managerial services would put the economy in the $L \in (K_2, K_2 + K_1)$ scenario, where (47) is downward sloping, and an even greater expansion would result in $L = K_2$.

6.1.3 CES Steady States

The solid dot on the $L = K_1 + K_2$ line in Figure 3 is the equilibrium for a Leontief economy in which the supply of managerial services is elastic and of the form $M(v)$, with $M(\infty)$ the same as it is in the inelastic economy.[20] The equilibrium managerial participation rate in this example is 97.5%, which puts the supply of managerial services close to its maximum. Figure 3 also shows the steady states for economies obtained by replacing $F(K, L) = \min\{K, L\}$ with CES technologies that have elasticities of substitu-

[20] This steady state is determined by $v = q_2 \mathrm{D}g(m_\infty)$, $M(v) = \beta L + m_\infty K_2$, and the present-value conditions (45), evaluated at $C = z_1 K_1 + z_2 K_2$, with $\delta K_1 = \theta K_2$.

tion equal to $\varepsilon = 0.3$ and $\varepsilon = 1$. The productivity ratio z_2/z_1 and the other parameters of these CES technologies are chosen to produce the exact same consumption-sector factor shares as in the Leontief steady state, at the input vector (K_1, K_2, L) implied by that Leontief steady state (see Proposition A1 for the construction). Because the steady state conditions (38)-(40) (and more specifically, (41)-(42)) depend on the technology $C(K_1, K_2, L)$ only via its factor shares, this construction implies that the steady state equilibrium does not change as one varies the elasticity of substitution ε of F. The curves labeled "present value" are versions of (41) and the curves labeled "market clearing" are versions of (42).

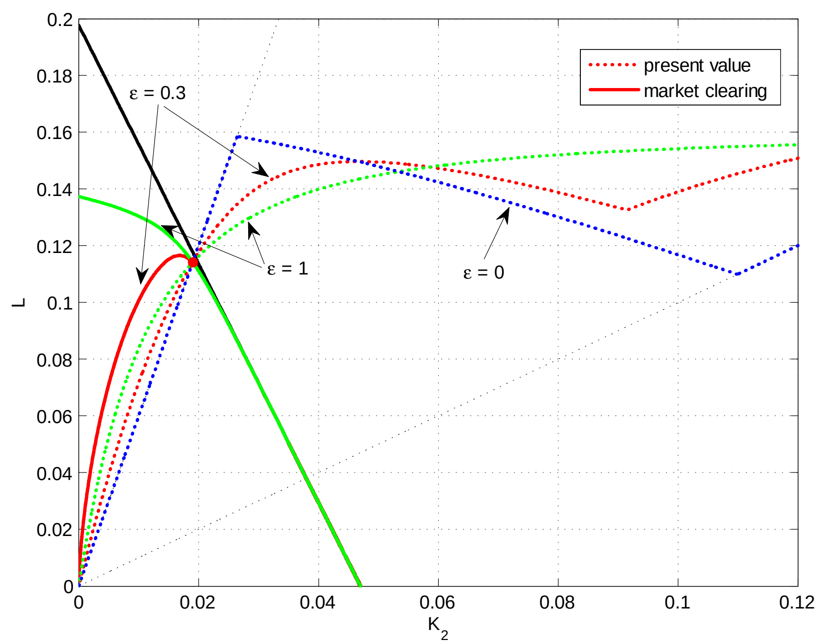

FIGURE 3 Zipf Steady States

For $\varepsilon = 0.3$, the present-value condition (41) is similar in shape to the Leontief curve (47): both are non-monotone and have a kink that arises because type-1 capital becomes obsolete when type-2 capital is abundant relative to the supplies of labor and managerial services. Obsolescence is not possible in the Cobb-Douglas case of $\varepsilon = 1$, and it is easy to show that (41) is then monotone increasing.

The market clearing condition (42) depends on the supply curve of managerial services. It constrains the set of feasible (K_2, L) to a strict subset of $\beta L + m_\infty K_2 \leq M(\infty)$ when the supply of managerial services is elastic. It is easy to show that (42) is still downward sloping when $\varepsilon = 1$. But it becomes hump-shaped when $\varepsilon = 0.3$. The downward sloping section of this hump shape is inherited from $\beta L + m_\infty K_2 \leq M(\infty)$, noting that a large K_2 can be accommodated only when $M(v)$ is close to $M(\infty)$, where the supply of managerial services is essentially inelastic. The upward-sloping section arises because the labor supply curve says that marginal utility weighted wages are high when L is high, and the technology implies that $\mathrm{D}_3 C(K_1, K_2, L)/C(K_1, K_2, L)$ is low. This squeezes the amount of team compensation that can go to managers when L is high, which lowers v and therefore $M(v)$ as well. This effect is strong when marginal products are very sensitive to inputs, as they are when $\varepsilon = 0.3$, but not when $\varepsilon = 1$. As $\varepsilon \downarrow 0$, the curve (42) becomes more and more like a tent with an apex at the Leontief steady state.

6.2 The Eigenvalue Structure

Consider Leontief steady state depicted in Figure 3 and replace the elastic supply of managerial services by an inelastic supply at $M > M(\infty)$. For sufficiently high M, the economy will have a steady state in which the supply of type-2 capital is so large that $C(K_1, K_2, L) = z_2 K_2$. Type-1 capital produces no profits and its price is zero. The condition $v \geq q_1 \mathrm{D}g(0)$ is met simply because $q_1 = 0$. Near this steady state, this economy behaves exactly like a one-type economy with a Leontief technology and a (particularly high) depreciation rate $\delta + \theta$. As argued at the beginning of Section 5, the type-2 capital stock then returns to the steady state at the speed $(1 + \beta/m_\infty)(\delta + \theta)\mathcal{S}_g$. Similarly, a sufficiently large M will make type-1 capital obsolete in the CES economy with $\varepsilon = 0.3$, and then the speed formula (33)-(34) for a one-type economy applies. But the economy now has two state variables. In the background, type-1 capital evolves according to $\mathrm{D}K_{1,t} = -\delta K_{1,t} + \theta K_{2,t}$, and this adds the eigenvalue $-\delta$.

Reducing M relative to this Leontief steady state will at some point put the economy into a steady state in which type-1 capital is also used to produce consumption but earns no profits—this corresponds to the downward-sloping part of (47). The price of type-1 capital is still zero, and small enough fluctuations in this type of capital will have no effect on equilibrium trajectories: the economy has enough "shovel ready" but low-quality capital. As a result, one can again apply convergence results for an equivalent one-type economy with a depreciation rate $\delta + \theta$, but now for one in which consumption equals

$z_1 L + (z_2 - z_1)K_2$. That is, the effective one-type technology is locally linear in capital and labor. Although this violates Assumption 2, the speed formula (33)-(34) applies, with a negative factor share elasticity equal to $\mathcal{E}_A = -(z_2 - z_1)K_2/(z_1 L + (z_2 - z_2)K_2)$. Again, the dynamics of $K_{1,t}$ implies an additional speed equal to δ.

6.2.1 Complex Eigenvalues

Further reducing M puts the economy in a regime in which the stock of type-2 capital that can be maintained is small, and so both high- and low-quality capital are fully employed to produce consumption. This is no longer an economy to which results for a one-type economy can be applied. But the Leontief technologies together with the Zipf limit imply that $\partial \mathrm{D}K_t/\partial q_t = 0$. In other words, the Jacobian for this economy is block diagonal. To see this, note that the Leontief technology forces $\beta L_t = \beta(K_{1,t} + K_{2,t})$, and then the Zipf limit implies a managerial resource constraint $M = \beta(K_{1,t} + K_{2,t}) + m_{2,t}K_{2,t}$. This mechanically determines $m_{2,t}$ as a function only of K_t, independently of q_t.[21] The resulting block diagonality means that the speed of convergence is determined by the eigenvalues of $\partial \mathrm{D}K_t/\partial K_t$ alone, as in the one-type scenario discussed at the beginning of Section 5. It is not difficult to verify that $\partial \mathrm{D}K_t/\partial K_t$ is given by

$$\begin{bmatrix} \frac{\partial \mathrm{D}K_{1,t}}{\partial K_{1,t}} & \frac{\partial \mathrm{D}K_{1,t}}{\partial K_{2,t}} \\ \frac{\partial \mathrm{D}K_{2,t}}{\partial K_{1,t}} & \frac{\partial \mathrm{D}K_{2,t}}{\partial K_{2,t}} \end{bmatrix} = \begin{bmatrix} -\delta & \theta \\ -\frac{\beta}{m_\infty} \times (\delta + \theta)\mathcal{S}_g & -\left(1 + \frac{\beta}{m_\infty}\right)(\delta + \theta)\mathcal{S}_g \end{bmatrix}.$$

The first row of this matrix is immediate from the Zipf dynamics $\mathrm{D}K_{1,t} = -\delta K_{1,t} + \theta K_{2,t}$. The non-zero slope $\partial \mathrm{D}K_{2,t}/\partial K_{1,t} = -(\beta/m_\infty)(\delta + \theta)\mathcal{S}_g$ arises from the fact that the low-quality capital stock $K_{1,t}$ is fully employed by teams of managers and workers, and so fluctuations in this type of capital will cause fluctuations in the residual supply of managerial services available for replicating type-2 capital. The slope $\partial \mathrm{D}K_{2,t}/\partial K_{2,t}$ includes the additional term $-(\delta + \theta)\mathcal{S}_g$ because type-2 capital is being replicated, and so fluctuations in type-2 capital also cause fluctuations in the $m_{2,t}K_{2,t}$ component of the demand for managerial services.

[21] Outside the Zipf limit, entry produces both types of capital, and replication can be directed. So the relative price $q_{1,t}/q_{2,t}$ will matter for $\mathrm{D}K_t$.

The eigenvalues of the matrix $\partial \mathrm{D}K_t/\partial K_t$ are

$$\lambda_{\pm} = -\frac{\delta + \left(1 + \frac{\beta}{m_\infty}\right)(\delta+\theta)\mathcal{S}_g}{2} \pm \sqrt{\left(\frac{\delta + \left(1 + \frac{\beta}{m_\infty}\right)(\delta+\theta)\mathcal{S}_g}{2}\right)^2 - \left(\frac{\delta}{\delta+\theta} + \frac{\beta}{m_\infty}\right)(\delta+\theta)^2\mathcal{S}_g}. \qquad (50)$$

Note that the argument of the square root in this expression is a quadratic function of the factor share $\mathcal{S}_g$. It is easy to verify that the $\lambda_{\pm}$ are real and negative for all $\mathcal{S}_g \in (0,1)$ close enough to the endpoints of $(0,1)$. Everywhere else (a non-empty subset of $(0,1)$), these eigenvalues are complex with a negative real part. The convergence to the steady state is then a damped oscillation—something that is not possible in a one-type economy. The resulting speed of convergence is $-\operatorname{Re}(\lambda_{\pm}) = (\delta + (1+\beta/m_\infty)(\delta+\theta)\mathcal{S}_g)/2$. This is simply the average of the depreciation rate of type-1 capital and the speed $(1+\beta/m_\infty)(\delta+\theta)\mathcal{S}_g$ that governs this economy in the $L = K_2$ scenario. If the $\lambda_{\pm}$ are real, then both eigenvalues are negative and the economy will have one speed that is slow relative to $(\delta + (1+\beta/m_\infty)(\delta+\theta)\mathcal{S}_g)/2$ and another that is fast.

6.2.2 CES $F(K,L)$ and Separable Factor Supplies

Figure 4 shows the half-lives for economies in which F is a CES production function with elasticities $\varepsilon \in [0,1]$, and in which the supplies of labor and managerial services are separable and elastic. To interpret and compare with the analytical results for the one-type economy and the two-type Leontief economy, it is necessary to specify the quantitative properties of these economies in some detail.[22]

Start with the $\varepsilon = 0.3$ benchmark economy. The subjective discount rate is $\rho = 0.04$ per annum, and organization capital is assumed to depreciate at a rate $\delta = 0.05$, with $\delta_k = 0.03$ and a firm exit rate at $\delta_f = 0.02$. This firm exit rate is much lower than the roughly 10% exit rate of employer firms in the US economy. But adding more randomness to post-entry growth can plausibly account for that (as in Luttmer [2007, 2011]). High-quality capital depreciates into low-quality capital at the rate $\theta = 0.25$. Because $g(0) - \delta_k = -0.03$, surviving low-quality firms shrink at the rate 0.03. Formula

[22] Only the key parameters and key steady state implications will be discussed here. A complete specification of functional forms and parameter values is given in Appendix D.

(13) for the tail index ζ then says that the right tail of the firm size distribution is determined by fast-growing firms with high-quality capital. The rest of the economy is specified so that $\zeta = 1.05$, close to what is found in US data (Luttmer [2007]). This entails $g(m_2) = 0.2871$, and so high-quality firms grow at an annual rate of almost 26%. Steady state capital stocks and prices are such that the implied aggregate economic depreciation rate is approximately 15%.

The technology for replication is CES with an elasticity of substitution equal to 0.6 and a steady state managerial factor share equal to $\mathcal{S}_g(m_2) = 0.6$. This generates a curvature $\mathcal{C}_g(m_2) = 2/3$, less than 1, but larger than the Cobb-Douglas value of $1 - \mathcal{S}_g(m_2)$. The entry technology is also CES, with the same steady state factor share but a much higher curvature, equal to $\mathcal{C}_f(n) = 2$. Recall from (9) that this combination of curvatures will temper fluctuations in entry rates relative to incumbent replication rates. New firms start with one unit of capital, and labor is measured in units that imply $L = K_1 + K_2$ in the steady state, consistent with $F(K, L) = \min\{K, L\}$ in the Leontief case. Given these units, the span of control parameter is $\beta/m_2 = 0.25$. In other words, it takes a four-unit reduction in labor to release enough managerial services to replicate one unit of high-quality capital at the equilibrium rate $g(m_2)$.

The fraction of households with an ability to supply only labor is 90%, and two-thirds of them are employed. The remaining 10% of households have the ability to supply managerial services, and 97.5% choose to do so in the steady state. This high participation rate is driven by average managerial earnings that are roughly 10 times those of workers. The distribution of managerial earnings is also more skewed than it is for workers. The resulting factor supply elasticities for labor and managerial services are $\mathcal{E}_L = 3$ and $\mathcal{E}_M = 0.1$, respectively.

But managers can and will switch between the supervision task and the entry and replication tasks. The residual supply elasticity $\mathcal{E}_{S,v}$ of managerial services available for entry and replication (that is, of $M(v) - \beta L(w)$) is determined by

$$\mathcal{E}_{S,v} = \frac{1}{1 - \frac{\beta L}{M}} \left(\mathcal{E}_M - \frac{\beta L}{M} \times \mathcal{E}_{w,v} \mathcal{E}_L \right),$$

where $\mathcal{E}_{w,v}$ is the elasticity of w with respect to v implied by (39) evaluated at $L = L(w)$.

Extending Lemma 2 slightly,

$$-\mathcal{E}_{w,v} = \frac{\frac{\beta v}{\beta v+w}}{\frac{w}{\beta v+w} + \left(\frac{\mathrm{D}_3 C(K_1,K_2,L)L}{C(K_1,K_2,L)} + \left(\frac{L_1}{L}\frac{1}{\mathcal{C}_1} + \frac{L_2}{L}\frac{1}{\mathcal{C}_2}\right)^{-1}\right)\mathcal{E}_L},$$

where $\mathcal{C}_j = (1-\mathcal{S}_j)/\varepsilon$ and $\mathcal{S}_j = \mathrm{D}_2 F(K_j, L_j)L_j/F(K_j, L_j)$ are the labor curvature and share parameters of F at (K_j, L_j) (compare to (34) and note that $1+\mathcal{E}_A = \mathcal{S}_j + \mathcal{C}_j$ in the one-type economy). The steady state delivers $\beta L/M \approx 0.6$ and $\beta v/(\beta v + w) \approx 0.44$. About 60% of managerial time is spent supervising workers, and managers account for as much as 44% of the cost of a team of managers and workers. The team factor share of $C(K_1, K_2, L)$ is 0.70 and the curvature of $C(K_1, K_2, \cdot)$ is approximately 0.67. The resulting elasticity $\mathcal{E}_{w,v}$ is only -0.10, implying much greater out-of-steady state variation in managerial wages than in worker wages. The net result is $\mathcal{E}_{S,v} \approx 0.68$, a relatively modest residual supply elasticity, but substantially greater than $\mathcal{E}_M = 0.1$.

The $\varepsilon = 0.3$ and $\mathcal{E}_L = 3$ benchmark is indicated with a solid dot in both panels of Figure 4. The low- and high-quality capital shares of $C(K_1, K_2, L)$ are 0.10 and 0.20, respectively, and this implies $z_1/z_2 \approx 0.47$ at $\varepsilon = 0.3$. The productivity ratio z_1/z_2 and the other parameters of F are adjusted to keep the factor shares of $C(K_1, K_2, L)$ constant at the $\varepsilon = 0.3$ steady state allocation (K_1, K_2, L). As in Figure 3, this ensures that the steady state does not change with ε. Similarly, the parameters of the labor supply curve are adjusted so that the $\mathcal{E}_L = 3$ benchmark steady state equilibrium value of (w, L) remains on the labor supply curve as $\mathcal{E}_L$ varies, without changing the participation rate of households who can supply labor. The Zipf economies in Figure 3 are just the large-population limits (Proposition 3) of the economies with elasticities $\varepsilon \in \{0, 0.3, 1\}$ used to generate the top panel of Figure 4. In the Zipf limit, m_2 increases toward a slightly higher $m_\infty = 1.0765$, and this results in a limiting factor share $\mathcal{S}_g = 0.5881$ and limiting ratio $\beta/m_\infty = 0.2377$. The resulting eigenvalues for the two-type Leontief economy are complex, with an implied half-life of 5.17 years.[23] At $\mathcal{S}_g = 0.6$ and $\beta/m_\infty = 0.25$, the two-type Leontief formula would predict 5.04 years, and the $\zeta = 1.05$ Leontief economy in Figure 4 has a half-life of 5.06 years. As expected, the Zipf approximation is innocuous.

As Figure 4 shows, increasing the CES elasticity ε above the Leontief value of $\varepsilon = 0$ causes the half-life of this economy to lengthen. This finding is in line with the role of

[23]The complex part of the eigenvalues is relatively small, resulting in a period of about 108 years. Practically, the economy will be at the steady state by the time a cycle is completed.

the curvature of F in the one-type economy (Section 5.3.2). At the benchmark value $\varepsilon = 0.3$ used to describe the Zipf steady state in Figure 3, the $\zeta = 1.05$ half-life is 6.16 years. For substitution elasticities slightly above $\varepsilon = 0.35$, the eigenvalues become real, with fast and slow speeds that diverge significantly as ε increases further. The fast and slow speeds of the Cobb-Douglas economy are 4 and almost 12 years, respectively.

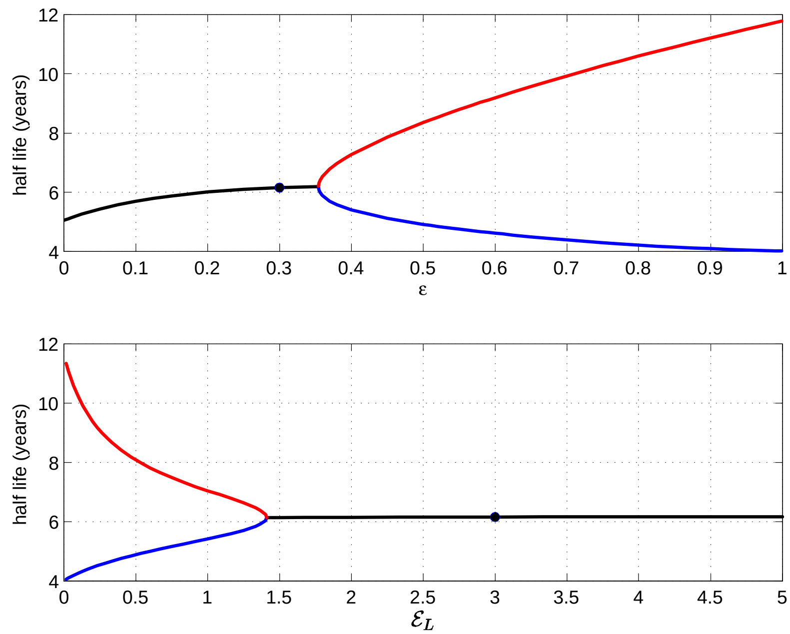

FIGURE 4 Eigenvalues and Elasticities

The bottom panel of Figure 4 shows how the half-lives of economies with a CES elasticity $\varepsilon = 0.3$ and a tail index $\zeta = 1.05$ vary with the labor supply elasticity $\mathcal{E}_L$. The figure shows that the half-life of this economy does not vary much with $\mathcal{E}_L$ for $\mathcal{E}_L$ above roughly 1.4. Below this threshold, the eigenvalues become real again. As the supply of labor becomes increasingly inelastic, the slow and fast speeds diverge, to values similar to those of the Cobb-Douglas economy with $\mathcal{E}_L = 3$.

For all economies displayed in Figure 4, low-quality capital accounts for 61% of consumption produced, and for 73% of team employment. Both the fast and slow eigenvalues will play a role in determining the rate at which aggregate consumption and employment converge to the steady state.

6.2.3 Speed Estimates from Misspecified Economies

Table 1 compares the speed implications for the two-type benchmark economy (with $\varepsilon = 0.3$ and $\mathcal{E}_L = 3$) with those inferred from speed formulas for simpler and therefore misspecified economies.

Data generated by the $\varepsilon = 0.3$ and $\mathcal{E}_L = 3$ economy produce a savings rate of 25.3%, a labor share (counting both managers and workers) of 67.4%, and an economic depreciation rate of 15.2%. Using these statistics together with a discount rate $\rho = 0.04$ and a labor supply elasticity $\mathcal{E}_L = 3$ to compute the speed of convergence of a Cass-Koopmans economy with a Cobb-Douglas technology gives 2.17 years—an application of the speed formula (36). This is the Cass-Koopmans number reported in Table 1.

TABLE 1

formula	half-life
Cass-Koopmans	2.17
one-type Leontief	3.17
one-type, $\mathcal{E}_{S,v} = 0.67$	5.26
one-type Cobb-Douglas	6.97
$(1 + \beta/m_\infty)\delta\mathcal{S}_g$, $\delta = 0.15$	6.16
two-type Leontief	5.17
exact at $\varepsilon = 0.3$	6.16

Clearly, this number severely underestimates the actual half-life of this economy. An alternative estimate that is easy to compute is the half-life of the one-type Leontief economy, based on (35) with the depreciated rate inferred from the equilibrium condition (32). The resulting depreciation rate is 14.5%, and then (35) implies a half-life of 3.17 years. This goes in the right direction but not by much. A more significant improvement results from adopting a CES specification and using $\mathcal{E}_{S,v}$ to infer $\mathcal{E}_A$ from (34). The one-type approximate speed (33) then implies a half-life of 5.26 years. This is still too low. But going all the way to a unit elasticity of substitution slows the economy down too much. Using $\mathcal{E}_A = 0$ and only the first factor in (33) results in a Cobb-Douglas half-life of 6.97 years.

A surprisingly accurate estimate comes from assuming a Leontief technology together with a completely inelastic supply of managerial services. As was argued in Section 5.1, the Zipf version of this economy implies the speed $(1+\beta/m_\infty)\delta\mathcal{S}_g$. Simply using $\beta/m_2 =$

0.25 and taking $\delta = 0.15$, very close to the economy depreciation rate of the benchmark two-type economy, matches the speed of that economy. This (almost) perfect estimate is also very different from the 3.17 years implied by the one-type Leontief economy with $\mathcal{E}_M = 0.1$, based on (35) with (32). The reason for these different estimates is not really $\mathcal{E}_M = 0.1$ versus $\mathcal{E}_M = 0$. This changes the estimated half-life by less than 3 months. The main reason is that (35) implicitly relies on $1 + \beta/m_\infty = 1/(1 - \beta L/M)$. The aggregate data from the two-type economy yield $\beta L/M \approx 0.6$ and thus $1/(1-\beta L/M) \approx 2.5$. From the perspective of a one-type economy, this implies $\beta/m_\infty \approx 1.5$. Used in the formula $(1 + \beta/m_\infty)\delta\mathcal{S}_g$ together with $\delta = 0.15$, this yields an estimated half-life of 3.08 years, not 6.16 years. The $\zeta > 1$ equilibrium of the two-type economy has $\beta/m_2 = 0.25$, and so $\beta/m_\infty \approx 1.5$ grossly overestimates this parameter and thereby the speed of convergence of this economy. Not all capital in the two-type economy is being replicated, and ignoring this heterogeneity leads to misleading estimates of the speed of convergence. One cannot use $M = \beta L + m_\infty K$ and $L = K$ with aggregate data to infer the key span of control parameter β/m_∞.

The last estimate reported in Table 1 is the two-type Leontief speed based on the real part of (50) with the actual depreciation rates δ and θ from the two-type economy. This estimate is very easy to calculate. But, as the top panel of Figure 4 shows, the speed of convergence is decreasing in the elasticity of substitution ε of F, and so this estimate is by construction an underestimate of the true speed for an economy with $\varepsilon = 0.3$.

6.3 Two Types of Shocks

The baseline two-type economy with $\varepsilon = 0.3$ and $\mathcal{E}_L = 3$ produces equilibrium trajectories that are not unlike what happens in a typical recession that is followed by a slow recovery. This will be illustrated here using two types of shocks. One is a one-time destruction of low-quality capital. In a richer model, some type of panic or temporary interruption of credit can cause low-quality firms to shut down; high-quality firms will find a way to survive. The second shock is a permanent decline in the rate at which consumers discount. Such a shock can mimic a news shock in which consumers learn that they are not as wealthy as they thought they were (see Luttmer [2013]). This shock puts the economy in a situation in which both low- and high-quality capital are below their respective steady states.

A Destruction of Low-Quality Capital Suppose the economy is in the steady state, and there is a one-time destruction of type-1 capital. Both types of capital are used to produce consumption, but type-1 capital is not replicated in the steady state. This continues to be the case after a destruction of type-1 capital that is not too large. Figure 5 shows the equilibrium trajectories of the quantities $K_{1,t}$, $K_{2,t}$, L_t, C_t, the factor prices measured in units of consumption v_tC_t and w_tC_t, and for the replication rate $g(m_{2,t})$ and the entry rate $f(n_t)$. For comparison with the $\varepsilon = 0.3$ benchmark, the trajectories for economies with $\varepsilon \in \{0, 1\}$ are also shown. The trajectories for the Zipf limit economies (not shown) are very close.

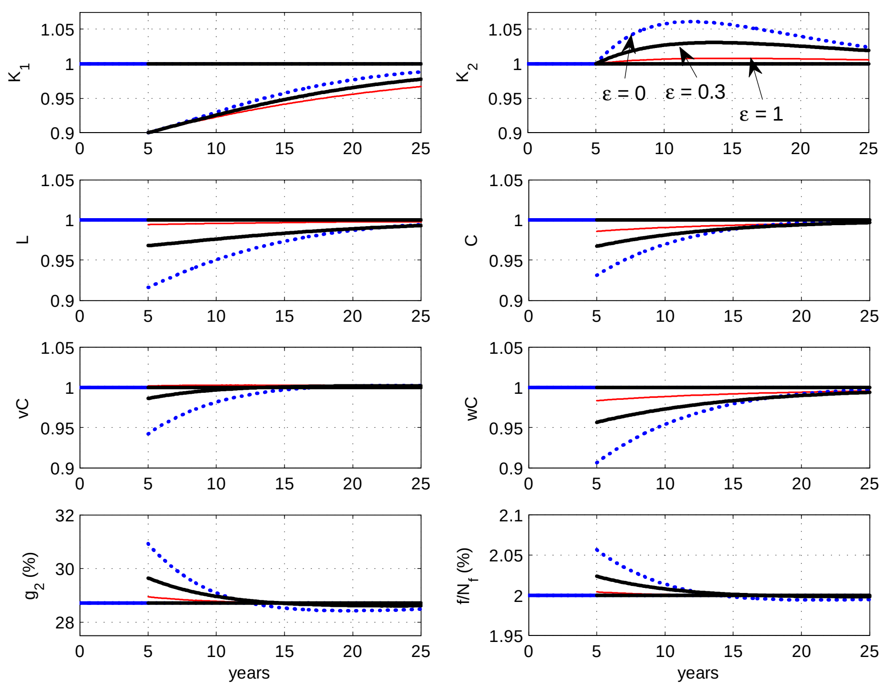

FIGURE 5 A 10% Destruction of Type-1 Capital

In the steady state, about 84% of the capital stock (measured so that new firms start with one unit of either high- or low-quality capital) is low-quality capital. In the Leontief economy, this means that a 10% reduction in K_1 results in a reduction in L that is almost as large. The decline in consumption is smaller because high-quality capital accounts for a disproportionate share of the aggregate output of consumption. Raising the elasticity of substitution to only 0.3 reduces the initial declines in L and C to about 3.2%. Because type-1 capital is not replicated, the stock of type-1 capital only recovers eventually because the (initially unaffected) type-2 capital stock rises temporarily above its steady state. On impact, managers released from overseeing workers in the production of consumption cause the rate at which the type-2 capital stock is replicated to rise from its steady state of 28.7% to about 31%.[24] Since the type-2 capital stock accounts for only about 16% of the steady state capital stock (27% of worker employment), it takes a long time for a significant recovery of worker employment to materialize. Because $\mathcal{C}_f = 2$, the response of entry is even more modest. But an entry technology with less curvature would not do much to speed things up because the economy is so close to the Zipf asymptote. The minimal impact of a 10% reduction in type-1 capital in the Cobb-Douglas economy is remarkable. As shown in Figure 4, the slow half-life for that economy is almost 12 years. If workers did not need supervision, then $\beta/m_2 = 0$ and the efficiency condition for worker employment would simply be $wL(w) = \alpha$. There would be no effect on worker employment at all. And Figure 5 shows that the $\beta/m_2 = 0.25$ economy remains quite close to that scenario.

A Discount Rate Shock

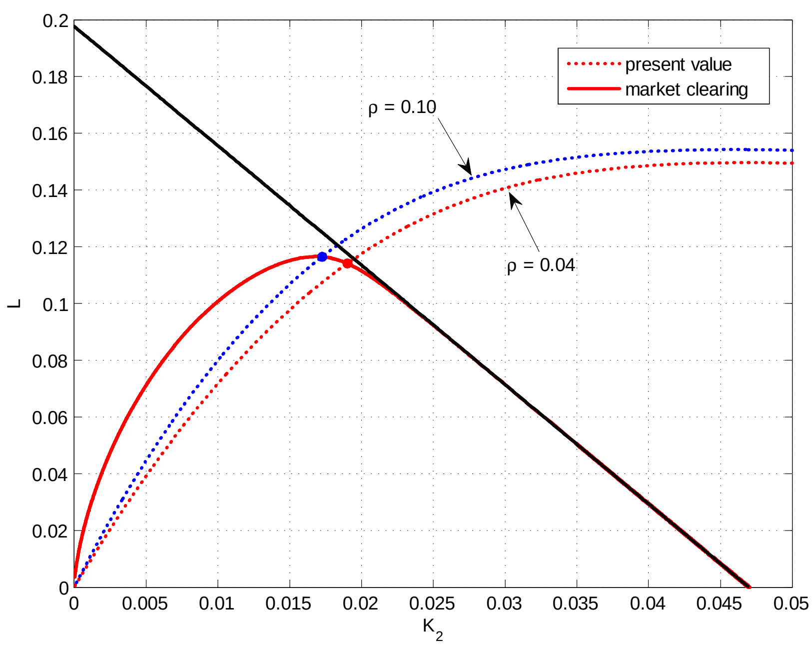

FIGURE 6 The Effect of ρ on the Zipf Steady State ($\varepsilon = 0.3$)

[24]Figure 1 in Luttmer [2012] presents some preliminary but suggestive evidence that the growth rates of fast-growing firms barely react to the state of the business cycle.

Suppose the economy is in the $\rho = 0.10$ steady state and there is an unforeseen once-and-for-all decline in ρ, to the benchmark value of $\rho = 0.04$. Figure 6 shows how the before and after Zipf steady states compare. As in Figure 3, the (mostly) upward-sloping dotted curve represents the efficiency condition (41), and the hump-shaped solid curve is (42)—effectively the resource constraint $M = \beta L + m_\infty K_2$, taking into account factor supply curves and the fact that v and w are inversely related via the equilibrium requirement that the cost $(\beta v + w)C$ of a team of managers and workers has to equal its marginal product. The first term in (41) is the cost share $\beta v/(\beta v + w)$ of managerial services in a team of managers and workers. Holding fixed capital-labor ratios, this cost share increases with a reduction in ρ because it raises present values and because of the first-order condition $v = q_2 \mathrm{D}g(m_\infty)$ that governs managerial incentives to replicate capital. In equilibrium, the cost share $w/(\beta v + w)$ equals $L^{-1}(L)C(K_1, K_2, L)/\mathrm{D}_3 C(K_1, K_2, L)$, and the only way this cost share can decline for a given (K_1, K_2) is through a reduction in L. Such a reduction also reduces profits per unit of capital, which helps to dampen the rise in present values. This explains the downward shift of (41) shown in Figure 6. Because this occurs in the region where (42) is downward sloping—a consequence of the fact that the managerial labor force participation rate is high—the permanent reduction in ρ causes steady state capital stocks to rise and worker employment to fall. In the benchmark economy with $\zeta > 1$, the decline in employment is actually a bit larger than in the Zipf limit shown in Figure 6.

The $\varepsilon = 0.3$ trajectories in Figure 7 show that worker employment overshoots in this economy: it declines by 4.7% on impact and by only 2% in the long run. Consumption declines by 3% on impact and rises by just under 2% in the long run.[25] Managerial and worker wages move in opposite directions as the economy shifts away from producing consumption and toward accumulating more organizational capital—following the same logic as the Stolper-Samuelson theorem of international trade. The alternative trajectories for $\varepsilon \in \{0, 1\}$ displayed in Figure 7 show that the choice of ε matters for the labor and consumption trajectories that follow a sudden decline in ρ. As long as both types of capital are fully employed in the Leontief economy, labor and consumption cannot decline on impact, and both will actually rise in the long run as the economy moves to

[25] From December 2007 to December 2009, the civilian employment-population ratio in the US fell by 4.4 percentage points, and real consumption per capita by about 3%.

the new steady state with higher stocks of both types of capital.[26] Worker employment does jump down in the Cobb-Douglas economy, but unlike in the $\varepsilon = 0.3$ economy, it barely recovers in the long run.

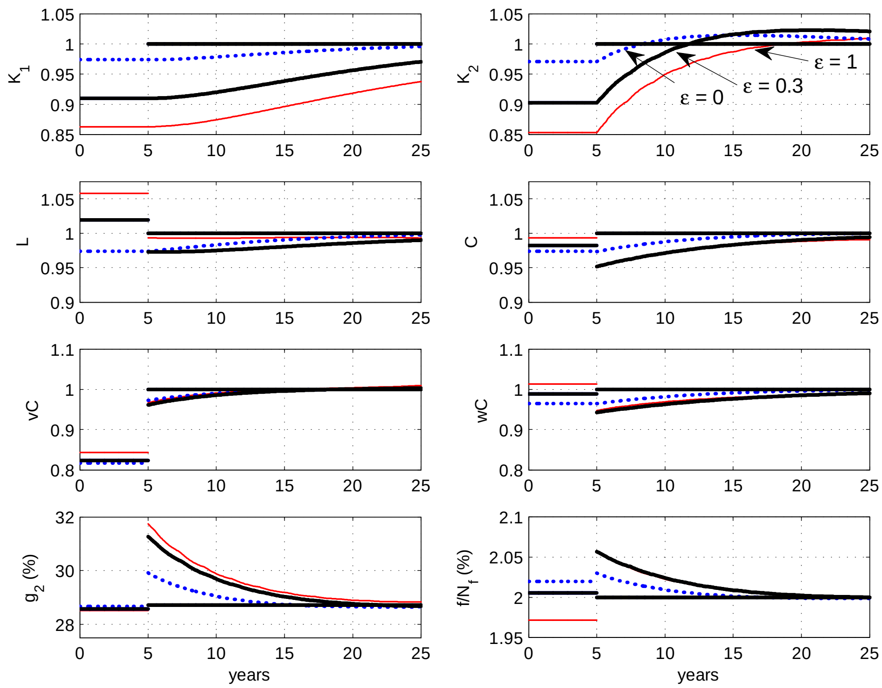

FIGURE 7 A One-Time Reduction in ρ

7. 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 to augment the capital stock is increasing and strictly convex in the rate at which capital is being accumulated. It is hard indeed 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.

[26] This all changes for Leontief steady states of the form $L \in (K_2, K_2 + K_1)$, which corresponds to the downward-sloping part of (47) in Figure 3.

The model presented here should be viewed as a proof of concept. It 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 in a manner that is consistent with some basic features of the business cycle. The results show that it is easy to construct model economies with half-lives that are similar to what has been observed following recent recessions. 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.

The results also show that it is hard to predict the true convergence speed of an economy on the basis of aggregate statistics alone. The underlying persistent heterogeneity in firm growth rates severely distorts the relation between aggregate share parameters and parameters that matter for the convergence rate of the economy.

A The Two-Type Firm Size Distribution

Consider the steady state and define $\mu_j = g(m_j) - \delta_k$. Write $z = \ln(k)$ for the "size" of a firm with k units of capital. For simplicity, normalize units so that new firms start with $k = 1$. Let $\tau \geq 0$ be the age of a firm at which it switches from high quality to low quality. The assumption is that $\tau = 0$ with probability ϕ_1 and $\tau > 0$ with probability ϕ_2. Given $\tau > 0$, the density of τ is $\theta e^{-\theta\tau}$. The random exit and quality transition events are assumed to be independent. Conditional on survival, the size of a firm is $z = Z(a, \tau)$, where

$$Z(a, \tau) = \mu_1 a + (\mu_2 - \mu_1) \min\{a, \tau\}.$$

In the following, begin by considering the case $\mu_{\mathrm{L}} > 0$ so that $Z(a, \tau) \geq 0$ for all firms.

The one-type calculation reported in the text immediately implies that $P[z|\tau = 0] = 1 - e^{-(\delta_f/\mu_1)z}$. Thus, δ_f/μ_1 is a possible tail index. Not surprisingly, the tail of $P[z|\tau > 0]$ may well decline more slowly toward zero. Consider a cohort of firms at age $a > 0$. The size of this cohort will be $e^{-\delta_f a}$. The independence of random exit and transition times implies that the density of transition times is still $\theta e^{-\theta\tau}$ among continuing firms. The distribution of size in the population of survivors at age a is therefore

$$\Pr\left[Z(a,\tau) \le z|a\right] = \int_0^\infty \iota\left[Z(a,\tau) \le z\right]\theta e^{-\theta\tau}\mathrm{d}\tau. \tag{51}$$

Observe that this implies $\Pr\left[Z(a,\tau) = \mu_2 a|a\right] = e^{-\theta a}$. Given $\tau < a$, the condition $Z(a,\tau) \le z$ corresponds to $\mu_1 a + (\mu_2 - \mu_1)\tau \le z$ and thus $\tau \le (z - \mu_1 a)/(\mu_2 - \mu_1)$. This is an empty event if $z < Z(a,0) = \mu_1 a$. Combining these observations with (51) gives

$$\Pr\left[Z(a,\tau) \le z|a\right] = \begin{cases} 0, & z \in [0, \mu_1 a), \\ 1 - e^{-\theta\left(\frac{z-\mu_1 a}{\mu_2-\mu_1}\right)}, & z \in [\mu_1 a, \mu_2 a), \\ 1, & z \in [\mu_2 a, \infty). \end{cases}$$

Aggregating this distribution over all age cohorts gives

$$\begin{aligned} P\left[z|\tau > 0\right] &= \int_0^\infty \Pr\left[Z(a,\tau) \le z|a\right]\delta_f e^{-\delta_f a}\mathrm{d}a \\ &= 1 - \frac{1}{\frac{\delta_f}{\mu_1} - \frac{\delta_f+\lambda}{\mu_2}}\left(\left(\frac{\delta_f}{\mu_1} - \frac{\delta_f}{\mu_2}\right)e^{-\left(\frac{\delta_f+\theta}{\mu_2}\right)z} - \left(\frac{\theta}{\mu_2}\right)e^{-\left(\frac{\delta_f}{\mu_1}\right)z}\right). \end{aligned}$$

The implied distribution of $k = e^z$ has the tail index reported in (13). Since the tail index of $P\left[z|\tau = 0\right]$ is $\delta_f/\mu_{\mathrm{L}}$, the overall tail index is the ζ reported in (13) for $\mu_1 = g(m_1)-\delta_k > 0$. If $\mu_1 < 0$, then slow-growing firms actually shrink, and they will certainly not appear in the right tail. It is not difficult to verify that then $\zeta = (\delta_f + \lambda)/\mu_2$.

B A Cobb-Douglas Lemma

As noted in the text, (16)-(17) imply that $\Sigma : \alpha \mapsto S(v)$ is a decreasing function. The proof of Lemma 4 relies on the observation that $L(v,w)$ is increasing α. This turns out to follow from the fact that the factor prices (v,w) must solve

$$\alpha = (\beta v + w)\,L(v,w), \quad \Sigma(\alpha) = M(v,w) - \beta L(v,w).$$

Differentiating this system with respect to α gives

$$\begin{bmatrix} \mathcal{E}_{L,v} & \mathcal{E}_{L,w} \end{bmatrix}\begin{bmatrix} \frac{\alpha}{v}\frac{\partial v}{\partial \alpha} \\ \frac{\alpha}{w}\frac{\partial w}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} \mathcal{E}_{L,v} & \mathcal{E}_{L,w} \end{bmatrix}\begin{bmatrix} \frac{\beta v}{\beta v+w} + \mathcal{E}_{L,v} & \frac{w}{\beta v+w} + \mathcal{E}_{L,w} \\ \frac{\mathcal{E}_{M,v} - \frac{\beta L}{M}\times\mathcal{E}_{L,v}}{1-\frac{\beta L}{M}} & \frac{\mathcal{E}_{M,w} - \frac{\beta L}{M}\times\mathcal{E}_{L,w}}{1-\frac{\beta L}{M}} \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ \frac{\mathrm{D}\Sigma(\alpha)\alpha}{\Sigma(\alpha)} \end{bmatrix}.$$

Using the fact that own price elasticities dominate cross price elasticities, one can verify that both the numerator and the denominator (the determinant of the matrix that must be inverted) are negative.

C The Cass-Koopmans Economy

Consider the same preferences as in the text, but now suppose that labor and managerial services are perfect substitutes in production. With slight abuse of notation, write $S(w) = L(w,w) + M(w,w)$. The labor market clearing condition is $m_t K_t = S(w_t)$, and output is given by $K_t g(m_t)$. This can be used as consumption or to add to the capital stock. Capital can also be consumed. The price of capital is therefore equal to 1, and so the marginal utility weighted price of capital is simply $q_t = 1/C_t$. This is also the marginal utility weighted price of output. The usual efficiency condition for labor is $q_t \mathrm{D}g(m_t) = w_t$. As in the text, the Euler condition is $r_t = \rho + \mathrm{D}C_t/C_t$, and hence $\mathrm{D}q_t/q_t = \rho - r_t$. The asset pricing equation for capital is $r_t = g(m_t) - \mathrm{D}g(m_t)m_t - \delta$.

The equilibrium conditions can therefore be summarized as

$$\begin{aligned} \mathrm{D}K_t &= (g(m_t) - \delta)K_t - \frac{1}{q_t} && (52) \\ \mathrm{D}q_t &= \left(\rho + \delta - [g(m_t) - \mathrm{D}g(m_t)m_t]\right) q_t && (53) \end{aligned}$$

together with the static equilibrium condition

$$m_t K_t = S(q_t \mathrm{D}g(m_t)). \tag{54}$$

Taking derivatives in (52)-(53) with respect to (K_t, q_t) and evaluating the result at the steady state gives

$$\begin{aligned} \begin{bmatrix} \frac{\partial \mathrm{D}K_t}{\partial K_t} & \frac{q_t}{K_t}\frac{\partial \mathrm{D}K_t}{\partial q_t} \\ \frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t} & \frac{\partial \mathrm{D}q_t}{\partial q_t} \end{bmatrix} &= \begin{bmatrix} -(\delta - g(m)) & -(\delta - g(m)) \\ \rho + \delta - g(m) & \rho + \delta - g(m) \end{bmatrix} \\ &+ \mathcal{S}_g g(m) \begin{bmatrix} \frac{K}{m}\frac{\partial m}{\partial K} & \frac{q}{m}\frac{\partial m}{\partial q} \\ 1 - \mathcal{C}_g \times \frac{K}{m}\frac{\partial m}{\partial K} & 1 - \mathcal{C}_g \times \frac{q}{m}\frac{\partial m}{\partial q} \end{bmatrix}. \end{aligned}$$

The second equation makes use of the steady state condition $1/(qK) = g(m) - \delta$, and the fourth equation relies on the steady state condition $\rho + \delta = (1 - \mathcal{S})g(m)$. The static equilibrium condition (54) is just like (8) but without entry. It follows almost immediately from (9) that

$$\begin{bmatrix} \frac{K}{m}\frac{\partial m}{\partial K} & \frac{q}{m}\frac{\partial m}{\partial q} \end{bmatrix} = \frac{1}{1 + \mathcal{C}_g \mathcal{E}} \begin{bmatrix} -1 & \mathcal{E} \end{bmatrix}.$$

Combining these results gives

$$\begin{bmatrix} \frac{\partial \mathrm{D}K_t}{\partial K_t} & \frac{q_t}{K_t}\frac{\partial \mathrm{D}K_t}{\partial q_t} \\ \frac{K_t}{q_t}\frac{\partial \mathrm{D}q_t}{\partial K_t} & \frac{\partial \mathrm{D}q_t}{\partial q_t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \rho & \rho \end{bmatrix} + \frac{g(m)\mathcal{S}_g}{1 + \mathcal{C}_g \mathcal{E}} \begin{bmatrix} -1 & \mathcal{E} \\ (1 + (1 + \mathcal{E})\mathcal{C}_g & 1 \end{bmatrix} + (g(m) - \delta) \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

This can be compared with (27).

D Calibration of the Two-Type Economy

The following describes the functional forms and parameter values used in the two-type economy of Section 6.

The number of employer firms in the US economy is about 6 million. For a potential workforce of about 200 million, this implies 3 firms per 100 households. Take $\delta_f = 0.02$, $\delta_k = 0.03$, and so $\delta = 0.05$. The resulting steady state flow of new firms is $f(n) = 6 \times 10^{-4}$ per household. Low-quality capital is not replicated, and so $m_1 = 0$ and $g(m_1) = 0$. The tail index (13) then becomes $\zeta = (\delta_f + \theta)/(g(m_2) - \delta_k)$. This is taken to be $\zeta = 1.05$, and $\theta = 0.25$. This implies $g(m_2) \approx 0.28714$. Among entrants, the fraction of fast-growing firms is $\phi_2 = 1 - \phi_1 = 0.4$. So quite a few new firms start by growing rapidly, but the mean duration of rapid growth is only $1/\theta = 4$ years. These parameters imply

$$K_2 = \frac{\phi_2 f(n)}{\delta + \theta - g(m_2)} \approx 0.0187, \quad K_1 = \frac{\phi_1 f(n) + \theta K_2}{\delta} \approx 0.1005.$$

The technology for producing new organization capital is determined by

$$\begin{aligned} f(n) &= \left((1-\omega_f) \left(\frac{1}{A_f} \right)^{1-1/\varepsilon_f} + \omega_f \left(\frac{n}{B_f} \right)^{1-1/\varepsilon_f} \right)^{1/(1-1/\varepsilon_f)}, \\ g(m) &= \left((1-\omega_g) \left(\frac{1}{A_g} \right)^{1-1/\varepsilon_g} + \omega_g \left(\frac{m}{B_g} \right)^{1-1/\varepsilon_g} \right)^{1/(1-1/\varepsilon_g)}. \end{aligned}$$

This parameterization requires one normalization each for f and g. Normalize $A_g = B_g$ and choose the units of managerial services so that $m_2 = 1$ in the steady state. Then $A_g = B_g = 1/g(m)$, and the steady state factor share of managerial services will be $\mathcal{S}_g(m_2) = \omega_g$. Also take $A_f = 1/f(n)$. This implies that $A_f n/B_f = 1$ and that the factor share of managerial services in the creation of new entrants is $\mathcal{S}_f(n) = \omega_f$. The steady state marginal products of f and g are now $\mathrm{D}f(n) = \omega_f/B_f$ and $\mathrm{D}g(m_2) = \omega_g/B_g$. The first-order condition $(\phi_1 q_1 + \phi_2 q_2)\mathrm{D}f(n) = q_2 \mathrm{D}g(m)$ therefore implies

$$\frac{B_f}{B_g} = \left(\phi_1 \times \frac{q_1}{q_2} + \phi_2 \right) \frac{\omega_f}{\omega_g}.$$

The equations for $q_j K_j$ reported in the text pin down q_1/q_2, and then B_f/B_g follows. The parameters used are $\omega_f = \omega_g = 0.6$, $\varepsilon_f = 0.20$, and $\varepsilon_g = 0.60$.

The production function $C(K_1, K_2, L)$ is defined by productivities z_1 and $z_2 > z_1$, and

$$F(K,L) = \left((1-\omega)\left(\frac{K}{A}\right)^{1-1/\varepsilon} + \omega\left(\frac{L}{B}\right)^{1-1/\varepsilon}\right)^{1/(1-1/\varepsilon)}.$$

This over-parameterization allows for a convenient normalization. The steady state conditions only depend on the factor shares of $C(K_1, K_2, L)$. The following proposition describes what these factor shares can tell us about F and the relative productivities of the two types of capital.

Proposition A1 *Suppose $C(K_1, K_2, L)$ has positive factor shares γ_1, $\gamma_2, 1-\gamma_1-\gamma_2$ at some $(K_1, K_2, L) \in \mathbb{R}^3_+$. Conjecture some $\varepsilon \in [0,1)$. Then the productivity ratio associated with the two types of capital must be*

$$\frac{z_2}{z_1} = \frac{\left(1-\gamma_1+\gamma_1\left(\frac{\gamma_2/K_2}{\gamma_1/K_1}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}{\left(1-\gamma_2+\gamma_2\left(\frac{\gamma_1/K_1}{\gamma_2/K_2}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}. \tag{55}$$

The parameter $\frac{\omega}{1-\omega}\left(\frac{A}{B}\right)^{1-1/\varepsilon}$ can also be identified from these factor shares. It is convenient to take $1-\omega = \gamma_1+\gamma_2$, and then it must be that

$$\frac{A}{B} = (\gamma_1+\gamma_2)\left(\frac{\gamma_1}{\gamma_1+\gamma_2}\left(\frac{K_1}{\gamma_1 L}\right)^{1-\varepsilon} + \frac{\gamma_2}{\gamma_1+\gamma_2}\left(\frac{K_2}{\gamma_2 L}\right)^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}.$$

The implicit allocation of labor is $L_j \propto K_j^{1-\varepsilon}\gamma_j^{\varepsilon}$, and the distribution of consumption produced is $C_j \propto (z_j K_j)^{1-\varepsilon}\gamma_j^{\varepsilon}$. For $\varepsilon = 1$, the ratios z_2/z_1 and A/B follow by taking the $\varepsilon \to 1$ limit in these expressions.

This follows from inverting standard first-order conditions. A corollary is that ε can be identified from the average capital productivities C_j/K_j and the type-j labor shares $\mathcal{S}_j = \mathrm{D}_2F(K_j, L_j)L_j/F(K_j, L_j)$, via

$$\frac{C_2/K_2}{C_1/K_1} = \left(\frac{\mathcal{S}_2}{\mathcal{S}_1}\right)^{\frac{1/\varepsilon}{1-1/\varepsilon}}\left(\frac{1-\mathcal{S}_2}{1+\mathcal{S}_1}\right)^{-\frac{1}{1-1/\varepsilon}}.$$

Such data would also deliver $z_2/z_1 = (\mathcal{S}_2/\mathcal{S}_1)^{(1/\varepsilon)/(1-1/\varepsilon)}$. The focus in this paper is on $\varepsilon \in (0,1)$. In these scenarios, high quality implies high average capital productivities and low labor shares.[27] The units of labor are chosen so that $L = K_1 + K_2$ in the steady state. In the case of $\varepsilon = 1$, this implies $A/B = 1$, and so $F(K, L) \propto \min\{K, L\}$. The factor shares of the two capital stocks are taken to be $\gamma_1 = 0.1$ and $\gamma_2 = 0.2$, resulting in a labor share at the conventional value of 0.7. At the benchmark value of $\varepsilon = 0.3$, this implies $z_1/z_2 = 0.467$, $C_1/(C_1 + C_2) = 0.5891$, and $L_1/(L_1 + L_2) = 0.6988$.

The distribution Ψ that defines the factor supplies (1)-(2) is taken to be a mixture of bivariate Fréchet distributions. A population of size H_A has $h_v = 0$ and (h_u, h_w) drawn from two independent Fréchet distributions with right tails that behave like $h^{-\sigma_A}$. A population of size $H_B = 1 - H_A$ has $h_w = 0$ and (h_u, h_v) drawn from two independent Fréchet distributions with right tails that behave like $h^{-\sigma_B}$. This gives rise to separable factor supplies of the form $L(w)$ and $M(v)$. They are the familiar logit supply curves

$$\begin{aligned} L(w) &= H_A \Gamma\left(1 - \frac{1}{\sigma_A}\right) [P_A(w)]^{1-1/\sigma_A}, \quad P_A(w) = \frac{(A_w w/A_u)^{\sigma_A}}{1 + (A_w w/A_u)^{\sigma_A}}, \\ M(v) &= H_B \Gamma\left(1 - \frac{1}{\sigma_B}\right) [P_B(v)]^{1-1/\sigma_B}, \quad P_B(v) = \frac{(B_v v/B_u)^{\sigma_B}}{1 + (B_v v/B_u)^{\sigma_B}}, \end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function. The population sizes are $H_A = 0.9$ and $H_B = 0.1$. The elasticity parameters are taken to be $\sigma_A = 10$ and $\sigma_B = 5$. The parameters (A_w, A_u) are set by matching the required steady state value for (w, L) and imposing an equilibrium worker participation rate $P_A(w) = 2/3$. Similarly, (B_u, B_v) is set to match the steady state value of (v, M) and $P_B(v) = 0.975$. The resulting factor supply elasticities are $\mathcal{E}_L = (1 - P_A(w))(\sigma_A - 1) = 3$ and $\mathcal{E}_M = (1 - P_B(v))(\sigma_B - 1) = 0.1$. Note that a high participation rate for type-B households tends to imply a small elasticity $\mathcal{E}_M$, unless σ_B is large. But if managers dominate the right tail of the earnings distribution, then σ_B cannot be large.

References

[1] Atkeson, A. and P.J. Kehoe, "Modeling and Measuring Organization Capital," *Journal of Political Economy*, vol. 113, no. 5 (2005), 1026-1053.

[2] Bachmann, R., "Understanding the Jobless Recoveries After 1991 and 2001," working paper, University of Michigan (2012).

[27] Given data generated by the two-type economies in this paper, Hsieh and Klenow [2009] would infer misallocation by imposing $\varepsilon = 1$ even though $\varepsilon \in (0, 1)$.

[3] Berger, D., "Countercyclical Restructuring and Jobless Recoveries," working paper, Northwestern University (2016).

[4] Bingham, N.H., C.M. Goldie, and J.L. Teugels, *Regular Variation*, Cambridge University Press (1987).

[5] Birch, D.L., *The Job Generation Process*, MIT Program on Neighborhood and Regional Change (1979).

[6] Cass, D. and K. Shell, "The Structure and Stability of Competitive Dynamical Systems," *Journal of Economic Theory*, vol. 12 (1976), 31-70.

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

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

[9] Granger, C.W.J., "Long Memory Relationships and the Aggregation of Dynamic Models," vol. 14, no. 2 (1980), 227-238.

[10] Greenwood, J., Z. Hercowitz, and G.W. Huffman, "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, vol. 78, no. 3 (1988), 402-417.

[11] Hopenhayn, H.A., "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, vol. 60, no. 5 (1992), 1127-1150.

[12] Hopenhayn, H.A. and R. Rogerson, "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, vol. 101, no. 5 (1993), 915-938.

[13] Hsieh, C.-T. and P.J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics,* vol. 124, no. 4 (2009), 1403-1448.

[14] Hurst, E. and B.W. Pugsley, "What Do Small Businesses Do?," *Brookings Papers on Economic Activity*, vol. 43, no. 2 (2011), 73-118.

[15] Jaimovich, N. and H.E. Siu, "Job Polarization and Jobless Recoveries," working paper, University of British Columbia (2015).

[16] 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.

[17] Koenders, K. and R. Rogerson, "Organizational Dynamics Over the Business Cycle: A View on Jobless Recoveries," *Federal Reserve Bank of St. Louis Review*, vol. 87, no. 4 (2005), 555-579.

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

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

[20] Luttmer, E.G.J., "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, vol. 2 (2010), 547-576.

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

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

[23] Luttmer, E.G.J., "The Stolper-Samuelson Effects of a Decline in Aggregate Consumption," *Federal Reserve Bank of Minneapolis Working Paper 703* (2013).

[24] Penrose, E., *The Theory of the Growth of the Firm*, John Wiley and Sons (1959).

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

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

[27] Rogerson, R., "Indivisible Labor, Lotteries and Equilibrium," *Journal of Monetary Economics*, vol. 21, no. 1 (1988), 3-16.

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

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

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