An Endogenously Derived AK-model of Economic Growth

Christian Jensen

University of South Carolina

February 2013

An Endogenously Derived AK-model of Economic Growth

Christian Jensen*
University of South Carolina

February, 2013

Abstract

Assuming a production process with returns to scale that vary with the intensity it is operated at, an AK-model of endogenous growth with constant returns to scale in production is shown to arise due to replication driven by profit-maximization. If replication occurs at the efficiency-maximizing scale, the result applies also when the number of production processes must be discrete, thus overcoming the so-called integer problem. When competition is imperfect, there is only convergence toward the AK-model for large enough input use, so an economy is more prone to stalling in a steady-state without growth, the smaller and less competitive it is. Inefficient scaling also raises the risk of stalling.

JEL Codes: O11; O40
Keywords: Economic growth; AK-model; Replication; Returns to scale in production; Integer problem

*Department of Economics; University of South Carolina; 1705 College Street; Columbia, SC 29208; 803-777-2786; cjensen@alumni.cmu.edu.

1 Introduction

In economic growth theory, the returns to scale in the factors of production that the economy accumulates endogenously are crucial for its long-run evolution. In particular, when these are decreasing, growth comes to a halt in the absence of any external impetus, as illustrated by Solow (1956). Therefore, models with sustained endogenous growth have increasing or constant returns to scale in the endogenously accumulated inputs, including those first proposed by Romer (1986) and Lucas (1988). Constant returns to scale are particularly popular, giving rise to the well-known AK-model (Rebelo (1991)), which can endogenously sustain a strictly positive constant rate of growth, a so-called balanced-growth rate, consistent with the empirical stylized facts first described by Kaldor (1963). Despite its crucial role, the nature of the returns to scale is always assumed, never derived, which is a major deficiency of endogenous-growth models (McCallum (1996)). At most, authors argue that replication leads to non-decreasing returns to scale, since if a production process can be reproduced exactly, the copy should arguably yield the same output as the original (Koopmans (1957), Shell (1966), Romer (1990, 1994) and Jones (1999, 2005)). However, this is insufficient, as it says nothing about production levels that are not a multiple of what the original process yields, and can therefore not be generated by copying at the original scale; the so-called integer problem (Romer (1990)).

If one assumes it is possible to copy the production process with the same degree of efficiency at any scale, returns to scale are constant for all levels of production, but by assumption. Instead, we assume that the efficiency of a production process, including its returns to scale, varies with the intensity it is operated at, an idea that goes back at least to Marshall (1890). Hence, exact replication, including the scale, yields the same output as the original, but copies that are scaled up or down are not equally efficient. Consequently, producers' scaling and replication of the process determines the returns to scale in production. In order to maximize efficiency, producers operate the process as close as possible to the intensity where its returns to scale are constant, and replicate it as production increases. However, they cannot do so exactly at this intensity, because the number of processes must be a positive integer, since otherwise the scale of the production process can be varied without affecting its efficiency, for example by running half a process at the efficiency-maximizing scale, thus violating our fundamental assumption. Only when the process is replicated at the efficiency-maximizing intensity does total production have constant returns to scale, which requires that nothing distort the profit-maximizing intensity away from the efficiency-maximizing one, and that competition be perfect, or with imperfect competition, that the number of processes, and thus replications, be large enough.

Because constant returns to scale, and the AK-model, only arise for large enough input use when competition is imperfect, endogenous growth can come to a halt in smaller economies, even when larger, but otherwise identical economies, keep on growing forever. Moreover, the marginal product of the input would be lower in the stalled economy, making it unable to attract a flow of input from the other. The reason is that it is harder to achieve the efficiency-maximizing scaling in smaller economies, which reduces the marginal products of inputs and affects their accumulation. As a result, an economy's starting point not only affects its growth rate, but could even determine whether it will stall or keep growing endogenously forever. Consequently, policies usually considered to have only a short-term impact on the rate of growth, such as a temporary inflow of inputs, or a transitory increase in competition, can have permanent effects, by getting a stalled economy on to the path of never-ending endogenous growth.

There are many reasons why the efficiency of a production process can vary with the scale it is operated at. One is the physical nature of the process, for example, in mineral extraction returns to scale might be decreasing as a result of the most easily extractable resources being exploited first. Another is specialization, which can make efficiency increase with the scale, as each worker concentrates more and more on the task at which he has a comparative advantage, instead of having to do a little of everything.[1] Returns to scale might initially be increasing due to the fact that it takes some time to get accustomed to performing a task, and doing so efficiently, while they turn decreasing as fatigue or boredom kicks in. Additional factors that can contribute to increasing returns to scale are fixed costs, synergies and learning-by-doing. Decreasing returns to scale can arise due to coordination and communication problems, which are more likely to emerge the larger the scale of operation.[2] It can also be harder to provide proper supervision and motivation in larger units, where the incentives to free-ride are greater.

Assuming a logarithmic production process with returns to scale that go from being increasing to decreasing as production rises, the next section derives an AK production function with constant returns to scale based on replication driven by producers'

[1] In order for the degree of specialization to vary with the scale, there must be indivisibilities in production (Edgeworth (1911), Kaldor (1934), Wicksell (1934) and Lerner (1944)).

[2] The managerial input can lead to decreasing returns to scale (Marshall (1890), Kaldor (1934) and Hicks (1939)) even when it increases proportionally with all other inputs, as it becomes overstretched due to the more than proportional complexity of the organization. The same applies for communication.

efforts to maximize efficiency. The following two sections define the rest of the general equilibrium model and study its dynamics, respectively. Subsequently, we show that when the number of processes must be a positive integer, there is only convergence toward constant returns to scale when replication occurs at the efficiency-maximizing intensity.

2 Production with optimal replication

Imagine the output y of a production process depends on the input k through the function

$$y = a\log(bk) \tag{1}$$

with given constants $a > 0$ and $b > 0$. Its returns to scale are increasing for $k \in (1/b, e/b)$, decreasing for $k > e/b$, and constant at $k = e/b$.[3] Output is zero for $k = 1/b$, and strictly negative for $k < 1/b$. Because the production process is a concave function, the optimal allocation among multiple identical processes is symmetrical, so the most total output Y that can be produced with N processes and K total input is

$$Y = Na\log\left(b\frac{K}{N}\right) \equiv H(K, N) \tag{2}$$

where $Y = N \times y$ and $K = N \times k$. If N could be varied continuously, the first-order condition

$$\frac{\partial H(K,N)}{\partial N} = a\left(\log\left(b\frac{K}{N}\right) - 1\right) = 0 \tag{3}$$

would yield the optimal number of processes

$$N^*(K) = be^{-1}K \tag{4}$$

since H is concave in N. The corresponding production function

$$Y = H(K, N^*(K)) = abe^{-1}K \equiv AK \tag{5}$$

yields the most output that could be produced with any amount of total input K. For given a and b, output Y would be linear in the input K, and have constant returns to scale, so the AK-model would arise for any N and K. However, if N could be varied continuously, it would be possible to change the scale of the production process

[3] Returns to scale at x_0 are said to be increasing when $f'(x_0) > f(x_0)/x_0$, decreasing when $f'(x_0) < f(x_0)/x_0$, and constant when $f'(x_0) = f(x_0)/x_0$, for any function $f(x)$.

without affecting its efficiency, for example by running half a process at the efficiency maximizing scale, thus violating our fundamental hypothesis, and making returns to scale constant by assumption.

When the number of production processes must be a positive integer,

$$Y = \max\left\{0, a\log\left(bK\right), 2a\log\left(b\frac{K}{2}\right), 3a\log\left(b\frac{K}{3}\right), \ldots\right\} \equiv F\left(K\right) \tag{6}$$

is the most output that can be produced with a given amount of input K. We imagine that the number of production processes must be varied in discrete units, while the input can be varied continuously. As an example, one can think of the number of workers and the hours that each works. It is not possible to hire half a worker, but it is possible to hire one to work part-time. The distinction is relevant when, as we assume, a worker's productivity depends on the number of hours worked. As one can easily imagine, the joint output of two part-timers working four hours each can differ from that of someone working the full eight hours. Just as one can hire someone to work eight hours a day, one can do so for eight hours and five minutes. Hence, while the adjustment on the extensive margin is restricted to integers, that on the intensive margin is not.

Figure 1, which plots $F(K)$, AK and $H(K,N)$ for $N = 1,2,3,4$ and $a = b = 1$, illustrates how $F(K)$ converges toward AK as K increases. Mathematically, the convergence can be shown as follows. For any K such that $N^*(K) = be^{-1}K$ is a positive integer, $F(K) = AK$. For any K such that $N^*(K)$ is not a positive integer, let

$$I\left(K\right) \equiv be^{-1}K - \lambda\left(K\right) \tag{7}$$

where $I(K) \in \mathbb{N}$ is the natural number closest to $be^{-1}K$ such that $I(K) < be^{-1}K$, implying that $\lambda(K) \in (0,1)$ and

$$F\left(K\right) = \max\left\{I\left(K\right)a\log\left(b\frac{K}{I\left(K\right)}\right), \left(I\left(K\right)+1\right)a\log\left(b\frac{K}{I\left(K\right)+1}\right)\right\}. \tag{8}$$

From the definition of $I(K)$ above (7), it follows that $K = (I(K) + \lambda(K))b^{-1}e$, so

$$F\left(K\right) = \max\left\{I\left(K\right)a\log\left(e + \frac{\lambda\left(K\right)}{I\left(K\right)}e\right), \left(I\left(K\right)+1\right)a\log\left(e + \frac{\lambda\left(K\right)-1}{I\left(K\right)+1}e\right)\right\} \tag{9}$$

while

$$AK = N^*\left(K\right)a\log\left(b\frac{K}{N^*\left(K\right)}\right) = \left(I\left(K\right)+\lambda\left(K\right)\right)a. \tag{10}$$

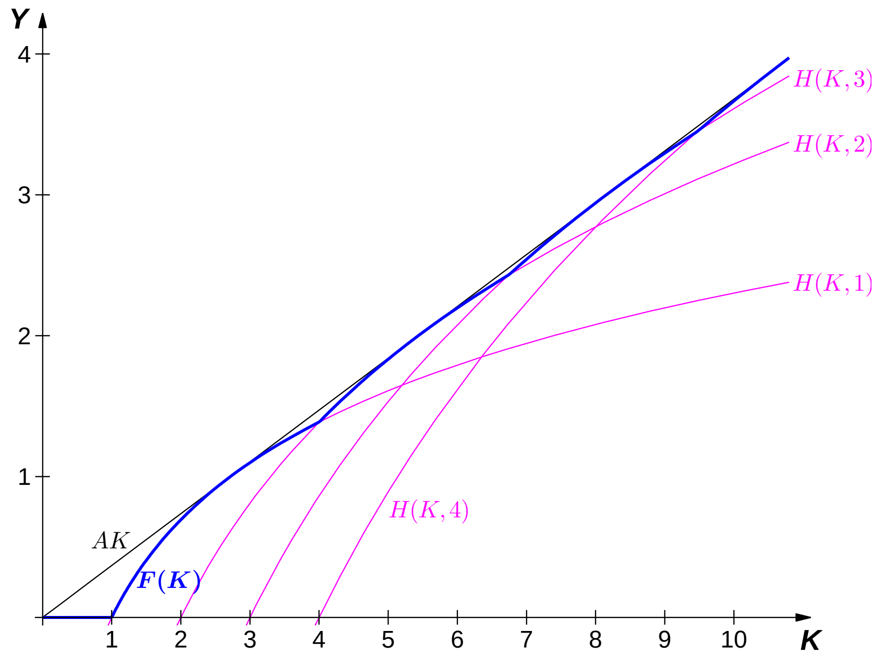

Figure 1: Convergence of $F(K)$ toward AK with efficient scaling.

According to Taylor's theorem (see for example Sydsæter et al. (1991)), for any function f that is twice continuously differentiable, there exists $\mu \in (0,1)$ such that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0 + \mu(x - x_0))(x - x_0)^2 \tag{11}$$

where the last element is Lagrange's error term for a first-order approximation of $f(x)$ around x_0. For any integer I, the function $I \log(K/I)$ is twice continuously differentiable with respect to K, so $\log(e + \lambda(K)/I(K)e)$ is too, and (setting $x = e + \lambda(K)/I(K)e$ and $x_0 = e$) we have

$$\log\left(e + \frac{\lambda(K)}{I(K)}e\right) = 1 + \frac{\lambda(K)}{I(K)} - \frac{\frac{(\lambda(K))^2}{2(I(K))^2}}{\left(1 + \phi\frac{\lambda(K)}{I(K)}\right)^2} \tag{12}$$

for some $\phi \in (0,1)$. Multiplying by $I(K)a$ yields

$$I(K)\, a \log\left(e + \frac{\lambda(K)}{I(K)}e\right) = I(K)\, a + \lambda(K)\, a - \frac{a\frac{(\lambda(K))^2}{2I(K)}}{\left(1 + \phi\frac{\lambda(K)}{I(K)}\right)^2} \tag{13}$$

so that

$$I(K)\,a\log\left(e+\frac{\lambda(K)}{I(K)}e\right)=AK-\frac{a\frac{(\lambda(K))^2}{2I(K)}}{\left(1+\phi\frac{\lambda(K)}{I(K)}\right)^2} \tag{14}$$

exploiting the decomposition above (10). Similarly, Taylor's theorem implies that

$$(I(K)+1)\,a\log\left(e+\frac{\lambda(K)-1}{I(K)+1}e\right)=AK-\frac{a\frac{(\lambda(K)-1)^2}{2(I(K)+1)}}{\left(1+\varphi\frac{\lambda(K)-1}{I(K)+1}\right)^2} \tag{15}$$

where $\varphi\in(0,1)$. As K grows, $N^*(K)=be^{-1}K$ increases, making $I(K)$ rise, and it follows from the expressions above (9, 14 and 15) that $F(K)$ converges toward AK. Hence, as replication increases with input use, returns to scale become constant, even if the number of production processes cannot be varied continuously.[4]

Producers seek to maximize profits, and producing efficiently, getting the most output possible from any amount of input, as determined above, is necessary for this. In addition, producers, who are assumed to be price-takers in the input market, must decide how much input to rent from consumers. Letting output be numeraire, having a price of one, and assuming perfect competition also in the output market, they do so by maximizing profits

$$\pi(K)=F(K)-(r+\delta)K \tag{16}$$

for a given rental rate r and depreciation rate $\delta\in(0,1)$, imagining input K is physical capital.[5] We must have

$$r=A-\delta \tag{17}$$

[4]If instead of being a constant, b were a non-rival input among production processes, so that $B=b$, the problem of choosing the optimal number of replications would remain unchanged and yield the production function $Y=ae^{-1}BK$, which has increasing returns to scale in inputs B and K jointly. Romer (1990 and 1994) and Jones (1999) argue that replication leads to increasing returns to scale, since technology, or ideas, are non-rival across production processes. Of course, not all innovations are non-rival, and even those that are non-rival are not always non-excludable. Schumpeter (1934), Griliches and Schmookler (1963), Schmookler (1966), Dasgupta and Stiglitz (1980), Grossman and Helpman (1991a, b), Rivera-Batiz and Romer (1991), Aghion and Howitt (1992), and Romer (1993) all stress the importance of profits, and thus excludability, in driving innovation. Because externalities in production are non-rival inputs, they too can affect the returns to scale, as is illustrated by Romer (1986) and Lucas (1988).

[5]The input does not have to be physical capital, but in order to complete the model, we need to take a stand on what it is and how it is accumulated. The input could even be a composite. For example, with the production process $y=\log(q^\alpha l^\beta)$, we have $Y=N\log((Q/N)^\alpha(L/N)^\beta)\equiv H(Q,L,N)$, which yields $N^*(Q,L)=e^{-1}Q^{\alpha/(\alpha+\beta)}L^{\beta/(\alpha+\beta)}$ and $Y=H(Q,L,N^*(Q,L))=(\alpha+\beta)e^{-1}Q^{\alpha/(\alpha+\beta)}L^{\beta/(\alpha+\beta)}$. Setting $a=\alpha+\beta$, $b=1$ and $k=q^{\alpha/(\alpha+\beta)}l^{\beta/(\alpha+\beta)}$ makes this framework identical to that above, with constant returns to scale in the composite input $K=Q^{\alpha/(\alpha+\beta)}L^{\beta/(\alpha+\beta)}$.

in a competitive equilibrium with nonnegative production. If the interest rate were higher than this, profits would be negative for all input levels, since the average product is at most A, so there would be no demand for input and no production. If the interest rate were lower than $A-\delta$, each producer would demand an infinite amount of input, as this would make its average product equal A, which would be greater than its average cost $r+\delta$, making profits infinitely large. Inserting for the equilibrium interest rate (17) into the profit function (16) yields

$$\pi(K) = F(K) - AK \leq 0 \tag{18}$$

which is strictly negative whenever F deviates from AK. As a result, a competitive equilibrium is only feasible when all production occurs at points on the production function where returns to scale are exactly constant.[6] Hence, when the economy-wide output changes, the number of producers and production processes adjusts so that all that remain active have an average productivity of A, since otherwise their profits would be strictly negative.

With imperfect competition in final goods, we have

$$r = \frac{\epsilon - 1}{\epsilon} F'(K) - \delta \tag{19}$$

where $\epsilon > 1$ is the elasticity of substitution between differentiated final goods (or the inverse of the elasticity of demand). This is the standard first-order condition for profit maximization with imperfect competition, after normalizing the price to unity, see for example Acemoglou (2008), and implies that producers apply a constant gross mark-up of $\epsilon/(\epsilon-1) > 1$ to the marginal cost of production $(r+\delta)/F'(K)$.[7] Each of the

$$H(K, I) = Ia \log\left(b\frac{K}{I}\right) \tag{20}$$

functions that make up $F(K)$, were $I \in \mathbb{N}_+$, is strictly concave in K for all $I > 0$. Hence, unless profits are maximized at a point where the number of production processes

[6]This is in line with Romer (1990, 1994) and Jones' (2005) point that perfect competition is incompatible with increasing returns to scale. In our setup it is also incompatible with K being at a point where $F(K)$ has decreasing returns, because a producer could then raise both her sales and profit margin with a large enough increase in input use. If the returns to scale of $F(K)$ were decreasing for all K, decreasing returns would be compatible with perfect competition.

[7]For example, with an inverse demand function $p(Y)$ for the output Y of a particular producer, yielding her relative price p as a function of her sales $Y = F(K)$, her profits are $p(F(K))F(K) - (r+\delta)K$. Assuming an interior solution, profit-maximization is given by the first-order condition $(\epsilon-1)/\epsilon \; p(F(K))F'(K) = r+\delta$, where $\epsilon = p'(F(K))F(K)/p(F(K))$.

changes, and we transition between $H(K,I)$ and $H(K+I)$, or at $K=0$, or for infinitely large K, the first order condition (19) must hold. The latter two are not feasible in an equilibrium with production, as they would imply a zero or infinite demand for input. To see that profits cannot be maximized at any point K_I where F transitions between $H(K,I)$ and $H(K+I)$, note that this would require

$$\frac{\epsilon-1}{\epsilon}H_1'(K_I,I)\geq r+\delta\geq\frac{\epsilon-1}{\epsilon}H_1'(K_I,I+1) \tag{21}$$

the first inequality so that profits do not rise as K is reduced along $H(K,I)$, the second so that they do not rise as K is increased along $H(K,I+1)$. The transition point K_I is given by $H(K,I)=H(K,I+1)$, which yields

$$K_I=\frac{(I+1)^{I+1}}{bI^I} \tag{22}$$

for $I=1,2,3,\ldots$ Exploiting that $H_1'(K,I)=IaK^{-1}$, we have

$$H_1'(K_I,I)=ab\frac{(I+1)^I}{I^I}<ab\frac{(I+1)^{I+1}}{I^{I-1}}=H_1'(K_I,I+1) \tag{23}$$

for finite I, contradicting the condition (21) necessary for profits to be maximized at K_I, $H_1'(K_I,I)\geq H_1'(K_I,I+1)$. The first-order condition (19) is also necessary for an equilibrium with production on the linear parts of F, which equal AK, since otherwise the marginal revenue $(\epsilon-1)/\epsilon A$ of using an additional unit of input would always be greater, or smaller, than the marginal cost $r+\delta$, making input demand infinitely large, or zero, respectively. The condition with perfect competition (17) is a special case of that with imperfect competition (19) for $\epsilon\to\infty$, so the more general expression is used below. Inserted into the profit function (16) it yields equilibrium profits

$$\pi=F\left(K\right)-\frac{\epsilon-1}{\epsilon}F'\left(K\right)K \tag{24}$$

which can be non-negative at points where F has increasing, decreasing or constant returns to scale.

The production function F assumes that the allocation among the underlying processes is optimal. Therefore, these must either be operated by the same producer, or be coordinated across producers in an effort to maximize efficiency and reduce costs. This can happen indirectly through outsourcing, or directly through arrangements such as code sharing among airlines. Alternatively, the coordination can arise as a result of all producers choosing their production levels so as to satisfy the same profit-maximizing first-order condition (19), assuming they have the same degree of market power. Hence, F can represent the production function of each individual producer, or that of the economy as a whole. In either case there is convergence toward constant returns to scale as the number of replications increases. With perfect competition, individual production satisfies constant returns to scale at all levels of replication, so the same applies to aggregate production.

3 Equilibrium

Consumers, who are assumed to be price-takers, rent out their capital S for a rate of return r. In addition, they collect profits π generated by production. These resources are used to accumulate capital and purchase consumption goods C. Consumption and saving decisions are made so as to maximize the discounted lifetime utility

$$\int_0^\infty \frac{C\left(t\right)^{1-\theta}-1}{1-\theta}e^{-\rho t}dt \tag{25}$$

subject to the budget constraint

$$\dot{S}\left(t\right)=r\left(t\right)S\left(t\right)+\pi(t)-C\left(t\right) \tag{26}$$

with respect to the control C and the state S, given paths for the interest rate r and profits π, and values for the constant relative risk-aversion parameter $\theta>0$, discount rate $\rho\in(0,1)$ and initial capital stock $S_0>0$. The first-order condition is

$$\frac{\dot{C}}{C}=\frac{r-\rho}{\theta} \tag{27}$$

the usual requirement for the optimal consumption path.[8]

The market-clearing condition for the input

$$K=S \tag{28}$$

determines the equilibrium rental rate r. Due to Walras' law, this condition also guarantees that the market for output clears. Combining the first-order conditions

[8] There is also a standard no-Ponzi game constraint, $\lim_{t\to\infty} S(t)\exp(-\int_0^t r(x)dx)\geq 0$, and transversality condition, $\lim_{t\to\infty} S(t)\exp(-\int_0^t r(x)dx)=0$, see Acemoglou (2008).

from maximizing profits (19) and utility (27) yields

$$\frac{\dot{C}}{C} = \frac{\frac{\epsilon-1}{\epsilon} F'(K) - \delta - \rho}{\theta} \tag{29}$$

while the budget constraint (26) becomes

$$\dot{K} = F(K) - \delta K - C \tag{30}$$

after substituting in for profits (16) and the market-clearing condition (28).

The production function $F(K)$ is not concave, so the first-order condition for the corresponding planner problem, $\dot{C}/C = (F'(K) - \delta - \rho)/\theta$, does not necessarily characterize the path that maximizes life-time utility (25) subject to the resource constraint (30), even when competition is perfect ($\epsilon \to \infty$). However, when households take the real interest rate and profits as given, their budget constraints (26) are linear, making the first-order condition (27) necessary and sufficient for optimality. The two solutions can differ because the planner might be willing to sacrifice current consumption to move to a point where the input is used more efficiently, thus allowing for higher future consumption. For individual households the rate of return, and efficiency with which the input is used, is given, so they cannot consider such trade-offs. By separating the decision of how much input to accumulate from that of how much to use in production, the non-concave problem is isolated to the simpler non-dynamic profit-maximization.

4 Dynamics

When competition is perfect, we have a standard AK-model for all levels of input, with consumption, production and capital always growing at the constant rate $\theta^{-1}(A-\delta-\rho)$ (see Rebelo (1991)). With imperfect competition, the dynamics are more complicated. Whenever $K \leq 1/b$, production $F(K)$ is, and always will be, zero, so both consumption and the stock of input approach zero. For $K > 1/b$, a steady-state equilibrium with constant non-negative consumption and input use exists for any capital $\bar{K}$ and consumption $\bar{C}$ satisfying

$$F'\left(\bar{K}\right) = I^*\left(\bar{K}\right) a\bar{K}^{-1} = (\delta + \rho)\frac{\epsilon}{\epsilon - 1} \equiv \kappa \tag{31}$$

and

$$\bar{C} = F\left(\bar{K}\right) - \delta\bar{K} = I^*\left(\bar{K}\right) a \log\left(b\frac{\bar{K}}{I^*\left(\bar{K}\right)}\right) - \delta\bar{K} \tag{32}$$

where $I^*(K) \in \mathbb{N}$ denotes the optimal discrete number of processes associated with input level K.

Figure 1 shows that before $F(K)$ converges to AK, its slope varies with K. F goes from one H-function to the next, so its slope $F'(K)$ is decreasing in K while moving along any one H-function, but jumps up each time $I^*(K)$ increases and F moves to a new H-function. Inserting for the transition points K_I and K_{I-1} from above (22) into $H_1'(K, I) = IaK^{-1}$, we find that while moving along $H(K, I)$, the slope

$$F'(K) \in \left(ab \left(\frac{I}{I+1} \right)^{I+1}, ab \left(\frac{I-1}{I} \right)^{I-1} \right) \tag{33}$$

for $I = 2, 3, 4, \ldots$ Along $H(K, 1)$, $F'(K)$ falls from ab to $.25ab$. As I increases, the lower bound for $F'(K)$ rises, while the upper bound falls, both converging toward $A = abe^{-1} \approx .368ab$.[9] Since the optimal number of processes $I^*(K)$ is increasing in K, it follows that an economy is more prone to getting stuck at a constant steady state the smaller K is. That is, when $\kappa \in (.25ab, .368ab)$, it is possible for an economy that starts out with little input to stall completely, while one that starts out with just a little more input could grow endogenously forever, even if the two economies were identical in all other respects. Moreover, the rate of return of the input would be lower in the stalled economy, preventing it from attracting input from the other.

If $\kappa < .25ab$ (and $K > 1/b$), consumption growth is always strictly positive, and therefore production and input use must also rise over time (though not necessarily in every period). If $\kappa > ab$, consumption is always shrinking, which can only be optimal if the economy itself is shrinking. If $\kappa \in (.25ab, .368ab)$, consumption growth can be positive or negative, but if the economy does not stagnate in a constant steady state and accumulates enough input, the consumption growth rate converges toward a strictly positive number. The closer κ is to .368ab, the greater input stock an economy can have and still risk stalling. If $\kappa \in [.368ab, ab)$, consumption growth can take any sign, but there is a limit to how much the economy can grow, since consumption growth would become negative, or zero, if it ever accumulated enough input for F to converge to AK. Whenever K is large enough for F to be indistinguishable from AK, we have a standard AK-model where consumption, input use and production are all growing at the constant rate $\theta^{-1}(A(\epsilon - 1)/\epsilon - \delta - \rho)$ (see Acemoglou (2008)).

The less competitive an economy is, the smaller is ϵ, and the higher is the threshold κ that the marginal product $F'(K)$ has to exceed in order to avoid stagnating in a steady-

[9]Convergence of $F'(K)$ toward A follows from $F(K)$ converging toward AK, but also from the fact that $\lim_{x \to \infty}(1 + m/x)^x = e^m$ for any constant m.

state without growth. Hence, according to our model, less competitive economies are more prone to stalling.[10] Of course, the degree of competition could easily change over time, thereby affecting the growth dynamics. For example, an economy stuck at a steady-state could start growing again endogenously if the degree of competition increased sufficiently. Moreover, it could keep growing endogenously forever, even if the increase in competition was just temporary.

5 Suboptimal replication

If the number of production processes could be varied continuously, returns to scale would be constant no matter the scale the process is replicated at. However, when the number of processes must be a positive integer, so that returns to scale are not constant by assumption, replication only leads to constant returns to scale with the efficiency-maximizing scaling. There are many circumstances that can distort a producers choice of how many processes to operate. Some examples are the time and costs incurred when setting up or dismantling a process, fixed costs associated with keeping it running, regulatory requirements that vary with the size of the operation, and credit constraints that inhibit producers from expanding at the efficiency-maximizing rate. For simplicity, we assume that the distortion is due to a government imposed tax or subsidy of τ per production process (transfered to the households). In addition, we imagine a perfectly competitive economy, and let $b = 1$. As a result, a producer's profits are given by

$$Na \log \frac{K}{N} - \tau N - (r + \delta) K \equiv G(K, N) \tag{34}$$

which is concave in N. Hence, the profit-maximizing number of processes is determined by the first-order condition

$$\frac{\partial G(K, N)}{\partial N} = a \left(\log \frac{K}{N} - 1 \right) - \tau = 0 \tag{35}$$

assuming the maximum is non-negative ($\tau < -a(1 + \log((r + \delta)/a))$. This yields the profit-maximizing number of processes

$$\hat{N}(K) = e^{-1-\frac{\tau}{a}} K \tag{36}$$

[10] The same applies for high depreciation and discount rates.

and the production function

$$Y = H\left(K, \hat{N}(K)\right) = (a+\tau)\, e^{-1-\frac{\tau}{a}} K \equiv \hat{A}K \tag{37}$$

which is linear, and thus has constant returns to scale. For $\tau \neq 0$, we have $\hat{A} < A$, reflecting that production is inefficient whenever the scaling is distorted.

When the number of processes must be a positive integer,

$$\max\left\{0, a\log K - \tau - (r+\delta)\,K, 2a\log\frac{K}{2} - 2\tau - (r+\delta)\,K, \ldots\right\} \tag{38}$$

yields the most profits that can be generated with a given amount of input K. The value of K at which one must switch from I to $I+1$ processes so as to maximize profits is now given by the point where $G(K,I)$ and $G(K,I+1)$ intersect, assuming profits are non-negative at such a point ($\tau < a(\log(aI/(r+\delta)\log(1+1/I)) - I\log(1+1/I))$). This yields the transition points

$$\hat{K}_I = \frac{(I+1)^{I+1}}{I^I} e^{\frac{\tau}{a}} \tag{39}$$

for $I = 1, 2, 3, \ldots$, which show that profit-maximizing producers use too few processes when $\tau > 0$, and too many when $\tau < 0$, compared to what maximizes output (22).

Figure 2 illustrates what happens when producers use too few processes. For $a = b = \tau = 1$, it plots $\hat{A}K$ and $H(K,I)$ for $I = 1, 2, 3, \ldots, 16$ (the latter are not labeled in the figure), together with the total production, labeled $\hat{F}(K)$, that results with the profit-maximizing transition points (39). These make production jump up as we go from one H-function to the next, because the transitions are not where the H-functions intersect (but instead where the G-functions intersect). One can easily show that

$$H\left(\hat{K}_{I+1}, I+1\right) - H\left(\hat{K}_I, I\right) = \tau \tag{40}$$

implying that production jumps by τ units whenever the number of processes increases by one. Hence, when replication happens at a suboptimal scale, production with a discrete number of processes does not converge toward a linear production function. Instead, it converges toward a piecewise linear function that jumps up or down each time the number of production processes changes. Because of these jumps, the production function does not satisfy constant returns to scale, even for large K.

Inserting for the transition points $\hat{K}_I$ and $\hat{K}_{I-1}$ into $H_1'(K,I) = IaK^{-1}$, we find

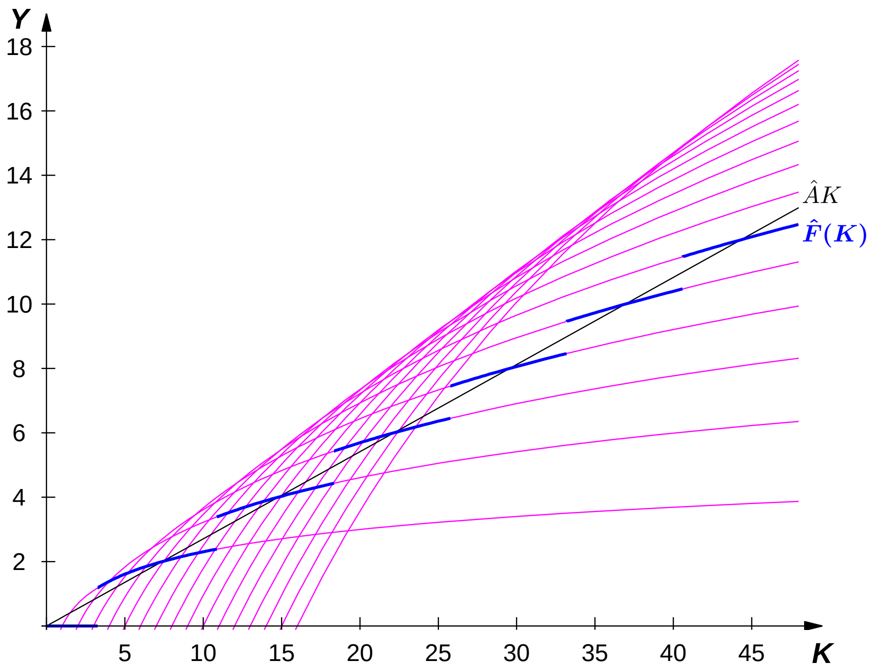

Figure 2: Lack of convergence toward AK with inefficient scaling.

that while moving along $H(K,I)$,

$$\hat{F}'(K) \in \left(a\left(\frac{I}{I+1}\right)^{I+1} e^{-\frac{\tau}{a}}, a\left(\frac{I-1}{I}\right)^{I-1} e^{-\frac{\tau}{a}} \right) \tag{41}$$

for $I = 2, 3, 4, \ldots$, which as I increases, converges towards $ae^{-1-\tau/a}$ ($\neq \hat{A}$). When $\tau > 0$, so that an inefficiently low number of processes is used, all operated at an inefficiently high scale, $\hat{F}'(K)$ is lower than it would be if efficiency were maximized, for all I, thus making the economy more prone to stalling in a steady-state without growth, or even shrinking over time. If $\tau < 0$, the number of processes is inefficiently high, the scale inefficiently low, and $\hat{F}'(K)$ is higher than it would be if efficiency were maximized, but the economy shrinks whenever the number of processes increases.

6 Conclusions

We show how the AK-model of economic growth can arise endogenously through the efficiency-maximizing replication of an underlying production process with returns to scale that vary with the intensity it is operated at. The result applies for a discrete number of replications, thus overcoming the so-called integer problem. When competition is perfect, the AK-model arises for all levels of input use and production. With imperfect competition, it only arises for a large enough number of replications, so it is possible for an economy to stagnate in a steady state without growth, while another that starts out with just a little more input, but is otherwise identical, could go on growing endogenously forever. Moreover, the marginal product of the input would be lower in the stalled economy, making it unable to attract a flow of input from the other. An economy is less prone to stalling the larger it is, and the higher the degree of competition among its producers. Our model suggests that even a temporary inflow of input, or transitory increase in competition, could start an everlasting growth spurt.

Even if replication leads to constant returns to scale in production, never-ending endogenous growth might not materialize. The reason is that the economy might not accumulate all inputs, making returns to scale decreasing in those it does accumulate endogenously. For example, in the Solow (1956) model, returns to scale are constant in capital and labor jointly, and therefore decreasing in capital alone, the only input it assumes that the economy amasses endogenously. Of course, as Lucas (1988) shows, what matters is not just what the economy accumulates in quantity, but also in quality. If it does not endogenously amass more workers, but does accrue human capital in terms of improved skills, it can still grow endogenously, even in per-capita terms. The same is true if land, usually considered to be available in given amounts for the economy as a whole, is used more intensively or efficiently. Because it is difficult to imagine an input that cannot be accumulated in quantity or quality, or used more efficiently, non-decreasing returns to scale all but guarantee sustained endogenous economic growth.

7 References

Acemoglu, D. (2008), *Introduction to Modern Economic Growth*, Princeton University Press.

Aghion, P. and Howitt, P. (1992), "A Model of Growth through Creative Destruction," *Econometrica* 60: 323-351.

Dasgupta P. and Stiglitz, J. (1980), "Uncertainty, Industrial Structure, and the Speed of R&D," *Bell Journal of Economics* 11: 1-28.

Edgeworth, F. Y., (1911), "Contributions to the Theory of Railway Rates," *Economic Journal* 21: 346-371 and 551-571.

Griliches, Z. and Schmookler (1963), "Inventing and Maximizing," *American Economic Review* 53: 725-729.

Grossman, G. M. and Helpman, E. (1991a), "Quality Ladders in the Theory of Growth," *Review of Economic Studies* 68: 43-61.

Grossman, G. M. and Helpman, E. (1991b), *Innovation and Growth in the Global Economy,* MIT Press.

Hicks, J. (1939), *Value and Capital: An inquiry into some fundamental principles of economic theory*, Clarendon Press.

Jones, C. I. (1999), "Growth: With or Without Scale Effects?" *American Economic Review* 89: 139-144.

Jones, C. I. (2005), "Growth and Ideas" in Aghion, A. and Durlauf, S. (Eds) *Handbook of Economic Growth*, Elsevier.

Kaldor, N. (1934), "The Equilibrium of the Firm," *Economic Journal* 44: 60-76.

Kaldor, N. (1963), "Capital Accumulation and Economic Growth" in Lutz F. A. and Hauge, D. C. (Eds) *Proceedings of a Conference Held By the International Economics Association*, Macmillian.

Koopmans, T. C. (1957), *Three Essays on the State of Economic Science*, McGraw-Hill.

Lerner, A. P. (1944), *The Economics of Control: Principles of Welfare Economics*, Macmillan.

Lucas, R. E. (1988), "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22: 3-42.

Marshall A. (1890), *Principles of Economics: An Introductory Volume*, Macmillan.

McCallum, B. T. (1996), "Neoclassical Vs. Endogenous Growth Analysis: An Overview," *Federal Reserve Bank of Richmond Economic Quarterly* 82: 41-71.

Rebelo, S. (1991), "Long-Run Policy Analysis and Long-Run Growth," *Journal of Political Economy* 99: 500-521.

Rivera-Batiz, L. A. and Romer, P. M. (1991), "Economic Integration and Endogenous Growth," *Quarterly Journal of Economics* 106: 531-555.

Romer, P. M. (1986), "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94: 1002-37.

Romer, P. M. (1990), "Endogenous Technological Change," *Journal of Political Economy* 90: S71-102.

Romer, P. M. (1993), "Idea Gaps and Object Gaps in Economic Development," *Journal of Monetary Economics* 32: 543-573.

Romer, P. M. (1994), "The Origins of Endogenous Growth," *Journal of Economic Perspectives* 8: 3-22.

Schmookler, J. (1966) *Invention and Economic Growth,* Harvard University Press.

Schumpeter, J. A. (1934), *The Theory of Economic Development*, Harvard University Press.

Shell, K. (1966), “Toward a Theory of Inventive Activity and Capital Accumulation,” *American Economic Review* 56: 62-68.

Solow, R. M. (1956), “A Contribution to the Theory of Economic Growth,” *Quarterly Journal of Economics* 70: 65-94.

Sydsæter, K., Strøm, A. and Berck, P. (1991), *Economists' Mathematical Manual*, Springer.

Wicksell, K. (1934), *Lectures on Political Economy I*, Routledge & Sons.