IJET International Journal of Economic Theory

doi: 10.1111/j.1742-7363.2009.00113.x

Entry and the accumulation of capital: A two state variable extension to the Ramsey model

Paulo Brito* and Huw Dixon†

In this paper we consider the entry and exit of firms in a dynamic general equilibrium model with capital. At the firm level, there is a fixed cost combined with increasing marginal cost, which gives a standard U-shaped cost curve with optimal firm size. Entry is determined by a free entry condition such that the cost of entry is equal to the present value of incumbent firms. The cost of entry (exit) depends on the flow of entry (exit). Equilibrium is saddle-point stable and the stable manifold is two-dimensional. Transitional dynamics can, under certain circumstances, be non-monotonic.

Key words entry, dynamics, Ramsey

JEL classification D92, C62, E32, O41

Accepted 19 September 2008

1 Introduction

The Marshallian approach to perfect competition focuses on a world where firms have U-shaped average cost curves and where the long-run competitive equilibrium is one in which all firms are producing at the efficient scale. As pointed out by Novshek and Sonnenschein (1987), this contrasts with the Arrow–Debreu setting where firms' production sets are convex, so that both average and marginal costs are non-decreasing.[1] Furthermore, the number of firms is taken as given in general equilibrium analysis

* ISEG (School of Economics and Management), Technical University of Lisbon and UECU (Research Unit on Complexity and Economics), Lisbon, Portugal.

† Cardiff Business School, Cardiff, UK. Email: dixonh@cardiff.ac.uk

This is a revised version of Brito and Dixon (2007). Financial support by FCT is gratefully acknowledged. This article is part of project POCTI/ECO/46580/2002, which was co-funded by the European Regional Development Fund. Huw Dixon would like to thank the National University of Singapore for its hospitality during the period of writing the final version of the paper. This paper was presented at the 1st Meeting of the *Portuguese Economic Journal* (University of the Azores, June 2007), Association of Southern European Economic Theorists Annual Meeting 2007 (Padua), Porto and Cardiff. We would like to thank the referee, Luís Costa, João Correia da Silva, Luigi Gandolfo, Aditya Goenka, Takashi Kamihigashi, Basant Kapur, Sheik Selim and Partha Sen for their comments on earlier drafts.

[1] McKenzie (1959) takes a quasi-Marshallian approach by assuming that an aggregate production function exists so that production sets are convex cones. This imposes the long-run constant returns of the Marshallian viewpoint by the argument that the possibility of division and replication implies that the aggregate production function will have constant returns.

(Smith 1974). Turning to explicitly dynamic macroeconomic models, the firm is often ignored unless there is imperfect competition. Where the number of firms is endogenous, although entry and exit are essentially dynamic phenomena, macroeconomists have modelled them as non-intertemporal phenomena. One approach is to have instantaneous free entry, so that the number of firms is that which ensures zero actual profits (see e.g. Devereux, Head, and Lapham 1996; Heijdra 1988; Coto-Martinez and Dixon 2003)[2] or zero expected profits (Hopenhayn 1992). An alternative is to treat the number of firms as fixed over time determined by a non-dynamic long-run zero profit condition (Hornstein 1993).

The present paper provides a rigorous treatment of the entry process in a perfectly competitive dynamic general equilibrium economy and traces the production process from the firm level to the aggregate. This model is a generalization of the classic Ramsey model in which the firm level is not modeled and the representative household chooses consumption as a control variable and where capital is the state variable. Here, there is an additional state variable (the number of firms) and control variable (entry). This results in a four-dimensional dynamic system, with a two-dimensional stable manifold. As we shall see, the related Ramsey model implicitly adopts two approaches: either there is a fixed number of firms or the number of firms adjusts instantaneously to the level of capital. However, these are too extreme: we allow for the case where the flow of entry is determined endogenously by an equilibrium entry model developed by Datta and Dixon (2002). The process of entry and the accumulation of capital interact in an explicitly dynamic setting. The entry model assumes that entry has a price at each instant in time, and this is increasing in the flow of entry and is zero when there is no entry (this is a special case of endogenous entry costs introduced by Das and Das (1996)). The dynamic equilibrium that results is one in which the cost of entry at each instant equals the net present value of an incumbent firm: as a consequence, firms have no incentive to delay or bring forward the time of entry/exit. In the steady state, there is zero entry, firms earn zero profits and we have the Marshallian long run, where average and marginal cost are equated. On trajectories towards the steady state, the flow of profits may be positive or negative and output per firm will differ from the efficient scale of production.

From the perspective of the representative household and the social planner, there are two ways of accumulating wealth: one is to set up new firms; the other is to accumulate capital. At all times, there is an arbitrage condition that ensures that the two assets have the same return. In the steady state the firms are at one level worthless, because they earn zero profits and have a zero marginal product. However, on a deeper level they are highly valuable: they enable the efficient organization of production in the steady state, where labor and capital are combined so that production occurs at on an efficient scale and marginal cost equals average cost. Even though firms earn no profits in the steady state, firms will be set up (or closed down) on the way to the steady-state equilibrium.

One of the most interesting findings of the present paper is that for a wide range of initial conditions we can have a non-monotonic trajectory in one of the state variables (but not both, because the roots are all real).[3] This occurs because of the interactions between

[2] This assumption is also made in the theory of contestable markets (see Baumol, Panzar, and Willig 1982).

[3] This complements findings in a non-general equilibrium context with non-monotonic entry dynamics (Gort and Klepper 1982; Das and Das 1996).

the two state variables: the number of firms influences the marginal product of capital, and the stock of capital influences the profitability of firms. For example, even if the number of firms is above the steady-state value, if there is a large capital stock this will boost firm profitability and lead to entry on the initial part of the trajectory. Likewise, a large number of firms boosts the marginal product of capital, which may lead to initial capital being accumulated even though in the long-run capital is decumulated.[4]

Existing papers that adopt a genuine dynamic entry model have assumed that there is only one state variable by removing capital. Aloi and Dixon (2002) adopt the same entry model as the present paper, but assume imperfect competition and labor as the only input. Bilbiie, Ghironi, and Melitz (2005) assume that there is an exogenous fixed entry cost and that the entrant evaluates the expected net present value of an incumbent. Free entry means that in equilibrium the net present value of an incumbent is equal to the entry cost. Again, this model assumes imperfect competition and no capital. Smith (1974) develops a genuine dynamic model of entry in a perfectly competitive economy. What determines the flow of entry in Smith's model is the opportunity cost of current consumption, because setting up a firm requires a one-off fixed labor input. Hence, if more firms are set up (or dismantled), then there is less labor available to provide for current consumption.

The paper is organized as follows. In Section 2 the model is presented. Section 3 studies the dynamics of the economy, and Section 4 concludes.

2 The model

There is an infinitely-lived household and at any time t a continuum of firms $i \in [0, n(t)]$. Households offer a fixed labor supply to firms and invest in their equity. Firms produce a single final product that is used for consumption and investment. Firms and households are price-takers in all markets. We now turn to the optimization programs of firms and households in more detail.

2.1 Household

Households consume and collect income from investments in financial assets (equity) and labor income. They choose the trajectory of consumption $\{C(t), t \geq 0\}$ to maximize lifetime utility U:

$$U = \int_0^{\infty} u(C(t))e^{-\rho t}dt,$$

where $u' > 0 > u''$, $u'(0) = +\infty; \lim_{C\to+\infty} u'(c) = 0$. The accumulation equation for financial assets is the instantaneous budget constraint

[4] In an earlier version of the paper, Brito and Dixon (2007), we studied two types of technology shocks: one is a productivity variable that alters the marginal product of labor and capital; the other alters the flow of fixed costs per firm. An increase in both the size and the number of firms occurs if the two shocks are positively correlated, and can result in non-monotonic trajectories in capital or in the number of firms.

$$\dot{V} = r(t)V(t) + w(t) - C(t), \tag{1}$$

where r and w are the real rate of return on equity and the real wage rate, respectively, and the labor supply is normalized to unity. The initial level of wealth $V(0)$ is given and the no-Ponzi-game condition $\lim_{t\to\infty} e^{-\int_0^t r(s)ds} V(t) = 0$ holds.

In the set of admissible consumption and wealth accumulation strategies, the optimal path of $(C(t),\ V(t))$ satisfies the Euler equation

$$\dot{C} = \frac{C}{\sigma(C)}(r(t) - \rho), \tag{2}$$

where $\sigma \equiv -u''(C)C/u'(C)$, and the transversality condition, which is

$$\lim_{t\to\infty} e^{-\rho t} u'(C(t))V(t) = 0. \tag{3}$$

2.2 Firms

There is a continuum of firms, $i \in [0,\ n(t)]$, where $n(t)$ is the measure of firms operating at instant t. Firms are price-takers in all the markets in which they participate: they hire labor and capital to produce output that can be consumed, invested, or used to set up (dismantle) firms. At every moment in time there is entry (or exit).[5] We define:

$$e(t) = \dot{n}. \tag{4}$$

2.2.1 Production

We start with the problem for incumbent firms $i \in [0,\ n(t)]$ in instant t. Each firm employs capital and labor according to the following technology:

$$y(i,\ t) = AF(k(i,\ t),\ l(i,\ t)) - \phi, \tag{5}$$

where $A > 0$ is a productivity parameter and $\phi > 0$ represents a fixed overhead in terms of final output. F is strictly concave, and homogeneous of degree $\nu < 1$ in capital and labor.[6] The Inada conditions hold for the marginal products of capital and labor. The average cost function corresponding to (5) is of the standard $U-$shaped variety: marginal cost is increasing because $\nu < 1$, and average cost is initially decreasing and then increasing because of the overhead element ($\phi > 0$). This implies that there is an optimal scale to the firm, where average cost is minimized. For any strictly positive $(w,\ r)$, average cost AC is minimized at the efficient firm size y^e:

5 That is, at instant t the number of firms will pass in the interval between t to $t+\epsilon$ from $n(t)$ to $n(t+\epsilon)$. If $n(t) < n(t+\epsilon)$ there is entry and if $n(t) > n(t+\epsilon)$ there is exit. The rate of entry is $\frac{n(t+\epsilon)-n(t)}{\epsilon}$. If the interval shrinks to zero, then the instantaneous rate of entry is $\dot{n}(t) = \lim_{\epsilon\to 0} \frac{n(t+\epsilon)-n(t)}{\epsilon}$.

6 Recall that homogeneity of degree ν for F implies the following relationships between $F(x,\ y)$ and its derivatives: $F_x x + F_y y = \nu F$, $(x,\ y)\cdot \frac{\partial^2 F(x,y)}{\partial(x,y)^2}\cdot(x,\ y) = \nu(\nu-1)F$, $xF_{xx} + yF_{xy} = (\nu-1)F_x$ and $xF_{xy} + yF_{yy} = (\nu-1)F_y$.

$$y^e = \frac{\phi \nu}{1-\nu}. \tag{6}$$

Note that A does not affect the efficient scale, although it does reduce optimal factor inputs. A decrease in ϕ reduces both efficient scale and factor inputs. As firms have the same technology, from now on we set $k(i, t) = k(t)$, $l(i, t) = l(t)$ and $y(i, t) = y(t)$ for any $i \in [0, n(t)]$. We define *firm size* by output $y(t)$.

We can define the supernormal profit of the firm π as the surplus when each factor is priced at its marginal product:

$$\pi(t) = y(t) - A(F_k k(t) + F_l l(t)) = (1-\nu) AF(t) - \phi. \tag{7}$$

The zero-profit condition is thus:

$$AF(k(t), l(t)) = \frac{\phi}{1-\nu}. \tag{8}$$

Note that this condition is equivalent to (6); hence, the zero profit condition implies technical efficiency when factors are priced at their marginal products (therefore, $P = MC = AC$). If output is above y^e, profits are strictly positive (because $P = MC > AC$) and negative if below y^e.

2.2.2 *Entry and exit*

The model of entry is based on Datta and Dixon (2002). Potential entrants (and quitters) evaluate the net present value, NPV, of incumbency,

$$NPV(t) = \int_t^\infty \pi(s) e^{-\int_t^s r(\tau)d\tau} ds. \tag{9}$$

We assume that there is a congestion effect that makes the marginal cost of entry (exit) q rise with the flow of entry (and exit); in particular, we assume that they are proportional:

$$q(t) = \gamma e(t), \tag{10}$$

which implies that the total cost of entry in terms of output used to set up (dismantle) firms is

$$Z(t) = \int_0^{|e(t)|} (\gamma i) di = \gamma \frac{e(t)^2}{2}, \tag{11}$$

where γ is a parameter measuring the dynamic barriers to entry (DBE). The congestion effect might arise from the setting up of firms at the same instant of time, stretching some finite resource: for example, if the technology for the setting up of new firms has diminishing returns.[7]

[7] A technology for producing (dismantling) firms of the form $\dot{n} = \sqrt{\frac{2Z}{\gamma}}$, where Z is the input invested into setting up (dismantling) firms, gives rise to (11).

The free entry condition means that the flow of entry equates the price of entry, q, to the net present value of incumbency:[8]

$$q(t) = \int_t^{\infty} \pi(s) e^{-\int_t^s r(\tau)d\tau} ds. \tag{12}$$

If we time-differentiate (12) we obtain

$$\dot{q} = -\pi(t) + r(t)q(t). \tag{13}$$

Rearranging (13), we can see that this is an arbitrage equation, equating the rate of returns on investment and setting up new firms:

$$r(t) = \frac{\pi(t)}{q(t)} + \frac{\dot{q}(t)}{q(t)}. \tag{14}$$

There are two elements to the right-hand side of (14): the profit earned by entering now and the change in the cost of entry. If entry costs are rising (falling) there is an incentive to bring forward (delay) the moment of entry. Substituting (10) into (13) we obtain the dynamic equation for entry:

$$\dot{e} = r(t)e(t) - \pi(t)\gamma. \tag{15}$$

The entry decision is inherently intertemporal. The entrant considers the future and decides whether or not to pay the entry cost now. An important implication of (9) is that entry can be nonzero when current profits are zero. Because it is the *NPV* of profits that matters, the entrant evaluates the flow of profits along the entire trajectory; therefore, for example, firms may enter even when current profits are negative if profits eventually become positive. As we shall see, this is exactly what happens along some equilibrium trajectories in this economy. This contrasts with the dynamic entry model of Howrey and Quandt (1968), where the flow of entry is related solely to the instantaneous profits.[9]

In equilibrium the entrant is indifferent between entering and not entering. Because this holds at each point in time, the potential entrant is also indifferent as to the timing of entry. For example, if the firm delays entering when the cost of entry is falling, it will find that the lower entry cost is exactly offset by the lower *NPV* of profits when it finally enters. This dynamic model of entry yields a dynamic zero-profit condition. The presence of entry costs means that the incumbents can earn strictly nonzero profits (losses) on the path to the steady state: the flow of entry adjusts so that the entry (exit) cost just balances the *NPV* of profits (losses) to be made. In the long-run steady state, the cost of entry is zero and both the *NPV* of incumbents and the flow of profits, π, are zero.

2.3 Aggregation

Let us denote the aggregate capital and labor available at time t as $K(t)$ and $L(t)$. For a given number of firms, the optimal allocation across firms is to have equal capital and labor

[8] Because the market for entry is competitive, the price of entry is equal to the marginal cost of entry.

[9] See also Meyers and Weintraub (1971) and Okuguchi (1972).

in each firm. This follows from the fact that marginal cost is everywhere increasing at the firm level. This is the outcome of decentralized factor markets where all firms face the same factor prices. Hence, $n(t)k(t) = K(t)$, $n(t)l(t) = 1$, so that the firm size and profits are:

$$y(t) = AF\left(\frac{K(t)}{n(t)}, \frac{1}{n(t)}\right) - \phi,$$
$$\pi(t) = (1-\nu)AF\left(\frac{K(t)}{n(t)}, \frac{1}{n(t)}\right) - \phi. \quad (16)$$

Aggregate output in the economy is $Y(t) = n(t)y(t)$, which from (16) we can write as a function of (K, n) and parameters (A, ϕ):

$$Y(K, n, A, \phi) = n\left[AF\left(\frac{K}{n}, \frac{1}{n}\right) - \phi\right]. \quad (17)$$

For analyzing the dynamics of the system, the marginal products of K and n, the two state variables are central. From (17) the aggregate and the firm-level marginal productivity of capital are equal, and there are decreasing returns at the aggregate level:

$$Y_K = AF_k > 0, \quad Y_{KK} = \frac{A}{n}F_{kk} < 0.$$

The derivative of aggregate output with respect to the number of firms is equal to the profit per firm,

$$Y_n = -A\left[F_k\frac{K}{n} + F_l\frac{1}{n}\right] + AF - \phi = (1-\nu)AF - \phi = \pi. \quad (18)$$

Hence, a zero profit equilibrium maximizes output $Y_n = 0$ if $\pi = 0$. Also, entry boosts the marginal product of capital (and labor):

$$Y_{Kn} = -\frac{A}{n}\left[F_{kk}\frac{K}{n} + F_{kl}\frac{1}{n}\right] = \frac{A}{n}(1-\nu)F_k > 0.$$

More firms means less inputs per firm so that the marginal products increase. Although there are diminishing returns at the firm level, the aggregate production function is homogeneous of degree 1 in (K, L, n): if you double capital and labor and the number of firms, productivity and output at the firm level are unaffected, so that aggregate output also doubles.

The real interest rate is the real rate of return on capital

$$r(K, n, A, \phi) = Y_K(K, n, A, \phi),$$

which is decreasing in the stock of capital and is increasing in the number of firms and the productivity parameter A, but is unaffected by the overhead: $r_K = Y_{KK} < 0$ and $r_n = Y_{Kn} > 0$, $r_A = Y_{KA} > 0$; $r_\phi = 0$.

From (18) we can write profits per firm as a function of aggregate variables:

$$\pi(K, n, A, \phi) = Y_n(K, n, A, \phi).$$

π is increasing in the stock of capital and the productivity and is decreasing in the number of firms and in the fixed cost: $\pi_K = Y_{Kn} > 0$, $\pi_n = Y_{nn} = -\nu(1-\nu)AF/n < 0$. By Young's Theorem, $\pi_K = r_n = Y_{Kn}$. Total operating profits in the economy are $\Pi(k, n, A, \phi) = n\pi$.

We have described a general firm-level technology in which there is a clearly defined optimal scale of production for firms that depends on the underlying technology and firms have the text-book cost curves with rising marginal costs and a U-shaped average cost curve. The role of firms is to define the way factor inputs are divided up and, hence, how efficiently they are combined. More firms means that capital and labor are divided up into smaller units, with the effect that their marginal products will increase but the additional fixed overheads might reduce or increase total output.

2.4 General equilibrium

The balance sheet for households equates the value of equity holdings to the total value of firms in terms of their assets (capital) plus the *NPV* of future profits:

$$V(t) = \int_0^{n(t)} [k(t) + q(t)]di = K(t) + n(t)q(t),$$

so that we can relate the change in the value of equity and capital accumulation as

$$\dot{K} = \dot{V} - n(t)\dot{q} - \dot{n}q(t) + Z(t), \tag{19}$$

where $\dot{n}q(t) - Z(t)$ reflects the impact of entry/exit on total equity value. From (4), (11), (13) and (19), we have the aggregate accumulation equation for capital:

$$\dot{K} = n\left[AF\left(\frac{K}{n}, \frac{1}{n}\right) - \phi\right] - C - \gamma\frac{e^2}{2}. \tag{20}$$

That is, output is equal to consumption plus investment in capital plus "investment," $Z(t)$, in setting up or dismantling firms.[10]

The household's transversality condition (3) can be related to $(K(t), n(t))$ by noting that $\lim_{t\to\infty} q(t) = 0$. Because $(K(t), n(t))$ are strictly positive it follows that:

$$\lim_{t\to\infty} e^{-\rho t}U'(C(t))n(t)q(t) = \lim_{t\to\infty} e^{-\rho t}U'(C(t))K(t) = 0. \tag{21}$$

Definition *The general equilibrium is defined by the paths of the aggregate variables* $(C(t), K(t), Y(t), V(t))$*, factor prices* $(r(t), w(t))$*, and the number and rate of change of firms* $(n(t), e(t))$*, for* $0 \le t < \infty$ *such that: (i) households determine* $C(t)$ *and* $V(t)$ *by maximizing lifetime utility subject to* (1, 3) *given the factor prices; (ii) incumbent firms*

[10] $Y(t) = C(t) + I(t) + Z(t)$, where $\dot{K} = I$, $Z(t) = \gamma e^2(t)/2$.

choose $(k(t), l(t), y(t))$ *to maximize profits, given factor prices; (iii) the flow of entry (exit)* $e(t) = \dot{n}(t)$ *equates the cost of entry (exit) with the NPV of an incumbent firm; (iv) trajectories* $(K(t), n(t))$ *satisfy the transversality condition (21); and (v) the factor prices ensure goods and factor markets clear.*

The paths of the aggregate variables $(C(t), K(t), e(t), n(t))$ are determined jointly from the four differential equations (2), (4), (15) and (20), the initial conditions $K(0)$ and $n(0)$ and the transversality conditions. From these we can determine firm-level variables $k(t)$ and $l(t)$, and firm size, $y(t)$.

2.5 Steady-state equilibrium

The steady-state aggregate capital stock and the number of firms, (K^*, n^*), will be determined by equating the marginal product of capital to the discount rate and setting profits equal to zero:

$$\rho = r(K^*, n^*, A), \tag{22}$$

$$0 = \pi(K^*, n^*, A, \phi). \tag{23}$$

Hence, there will be no entry and aggregate consumption, C^*, will be equal to aggregate output

$$e^* = 0, \tag{24}$$

$$C^* = Y^* = n^* y^*. \tag{25}$$

Because profits are zero, firm size is at the efficient level, $y^* = y^e$. Given the Inada property of the production function the steady state exists and, given global concavity, is unique.

Observe that the steady state depends on the technology parameters (A, ϕ) and the rate of time preference, ρ, but not on the elasticity of intertemporal substitution (σ) and DBE (γ). Changes in the last two parameters only influence the adjustment dynamics because they determine intertemporal arbitrage (postponement of consumption and of entry). The equations are recursive: the first two equations can be solved for (K, n), which then determines consumption by (25). We can make a geometrical projection of the phase diagram into (K, n)-space, as in Figure 1.

The two steady-state equilibrium conditions (22, 23) are invariant to the control variables (C, e) and the parameter γ. Their slopes are

$$\left.\frac{dn}{dK}\right|_{r=\rho} = -\frac{r_K}{r_n} > 0, \qquad \left.\frac{dn}{dK}\right|_{\pi=0} = -\frac{\pi_K}{\pi_n} > 0. \tag{26}$$

That is, they are both upward sloping but the $r = \rho$ line is steeper.[11] These two curves act as fixed reference points to which we can relate the elements defining the trajectories

[11] Observe that $\frac{\pi_n r_K - \pi_K r_n}{r_n \pi_n} = -\frac{F_{kk}F_{ll} - F_{kl}^2}{\nu(1-\nu)^2 F F_k} < 0$.

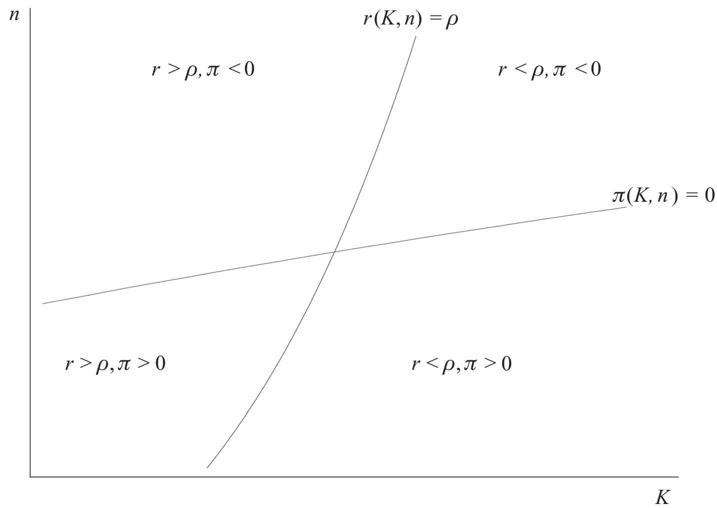

Figure 1 Steady-state equilibrium in (K, n).

projected in (K, n). They are also coincident with the projections of the isoclines $\dot{C} = 0$ and $\dot{e} = 0$ if all the variables are at their steady-state levels.[12] In Figure 1, to the left of the $r = \rho$ line $r > \rho$ and, hence, $\dot{C} > 0$; to the right $r < \rho$ and $\dot{C} < 0$. In addition, above the $\pi = 0$ line we have $\pi < 0$ and below it $\pi > 0$. Recall from (9) that we cannot infer the flow of entry from the instantaneous flow of profits, π, because entry/exit is determined by the *NPV* for the subsequent trajectory.

3 Aggregate dynamics

Next, we will characterize qualitatively the local dynamics properties in a neighbourhood of the steady state, by studying the solution of the linearized system

$$\begin{pmatrix} \dot{C} \\ \dot{e} \\ \dot{K} \\ \dot{n} \end{pmatrix} = \begin{pmatrix} 0 & 0 & Cr_K/\sigma & Cr_n/\sigma \\ 0 & \rho & \pi_K/\gamma & \pi_n/\gamma \\ -1 & 0 & \rho & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} C(t) - C^* \\ e(t) \\ K(t) - K^* \\ n(t) - n^* \end{pmatrix} \tag{27}$$

given the initial conditions $(n(0), K(0))$ and the transversality conditions (21), where the Jacobian matrix is denoted by J (henceforth the terms Cr_K and Cr_n in J are shorthand for $C^*_{r_K}$ and $C^*_{r_n}$).

3.1 The stable manifold

Our first step is to determine the eigenvalues.

[12] These isoclines will "slide" in (K, n) space when variables deviate from their steady-state values.

Proposition 1 *The eigenvalues of the Jacobian matrix of system* (27) *are given by*

$$\lambda_1^j = \frac{\rho}{2} \mp \left[\left(\frac{\rho}{2}\right)^2 - \frac{1}{2}\left(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma}\right) - \frac{1}{2}\Delta^{\frac{1}{2}} \right]^{\frac{1}{2}} \qquad j = s, u \tag{28}$$

$$\lambda_2^j = \frac{\rho}{2} \mp \left[\left(\frac{\rho}{2}\right)^2 - \frac{1}{2}\left(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma}\right) + \frac{1}{2}\Delta^{\frac{1}{2}} \right]^{\frac{1}{2}} \qquad j = s, u, \tag{29}$$

where the discriminant is $\Delta = (\frac{Cr_K}{\sigma} - \frac{\pi_n}{\gamma})^2 + 4\frac{Cr_n}{\sigma}\frac{\pi_K}{\gamma} > 0$, *and the superscripts s and u refer to stable and unstable eigenvalues, respectively. Hence, we have:*

$$\lambda_2^s < \lambda_1^s < 0 < \lambda_1^u < \lambda_2^u. \tag{30}$$

All proofs are in the Appendix. Note that $\lambda_1^s + \lambda_1^u = \lambda_2^s + \lambda_2^u = \rho$. It is sometimes useful to deal with the products of the eigenvalues: let $l_i \equiv \lambda_i^s \lambda_i^u$ for $i = 1, 2$, then

$$l_1 = \frac{1}{2}\left(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma} + \Delta^{\frac{1}{2}}\right), \quad l_2 = \frac{1}{2}\left(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma} - \Delta^{\frac{1}{2}}\right), \tag{31}$$

then $0 > l_1 > l_2$, $l_1 + l_2 = \frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma} < 0$ and $l_1 l_2 = \det(J) > 0$.

We have four distinct real eigenvalues: the two negative eigenvalues $\lambda_2^s < \lambda_1^s < 0$ will determine the dynamics for the transversality condition to hold. In our model, although the dynamics are saddle-path stable, the stable manifold is two-dimensional. Because the dimension of the stable manifold equals the number of predetermined variables the equilibrium is determinate.

The meaning of a two-dimensional stable manifold is that there are two independent sources of stability. If we look at the terms in the formula for the eigenvalues, we see parameters (ρ, σ), which reflect inter-temporal consumer preferences, γ, which relates the flow of entry to the cost of entry, and (r_K, r_n, π_n), the effect of the state variables on the marginal products (which depend only on technological parameters). The diminishing marginal returns of capital and decreasing profits to entry are the main forces for stability in this economy. It should be noted that they also interact: an increase in the number of firms increases the marginal productivity of capital; more firms means less capital per firm and, hence, a higher marginal product because $\nu < 1$. An increase in the capital stock makes firms more profitable, which affects entry. As we shall see, these interactions can lead to non-monotonic trajectories.

In the Appendix (proof of Proposition 2), we show that the dynamics of the system along the linearized stable manifold take the form:

$$\begin{pmatrix} C(t) - C^* \\ e(t) - 0 \\ K(t) - K^* \\ n(t) - n^* \end{pmatrix} = P_1^s w_1^s e^{\lambda_1^s t} + P_2^s w_2^s e^{\lambda_2^s t}, \tag{32}$$

where w_i^s are the weights (determined by the initial conditions for the state variables, $K(0)$ and $n(0)$), and P_i^s are the eigenvectors associated to the stable eigenvalues, $\lambda_i^s < 0$.

The space tangent to the stable manifold is a two-dimensional linear space E^s in (C, e, K, n) given by

$$E^s \equiv \{(C, e, K, n) : C - C^* = h_C(K - K^*, n - n^*), e = h_e(K - K^*, n - n^*)\}, \quad (33)$$

where h_C and h_e are linear functions,[13]

$$h_C = \left(\frac{\lambda_2^u - \lambda_1^u}{l_2 - l_1}\right)\left[\left(\frac{Cr_K}{\sigma} - \lambda_1^u\lambda_2^u\right)(K - K^*) + \frac{Cr_n}{\sigma}(n - n^*)\right], \quad (34a)$$

$$h_e = \left(\frac{\lambda_2^s - \lambda_1^s}{l_2 - l_1}\right)\left[\left(\frac{\pi_n}{\gamma} - \lambda_1^s\lambda_2^s\right)(K - K^*) + \frac{\pi_K}{\gamma}(n - n^*)\right], \quad (34b)$$

where the coefficients for K are both positive and for n are both negative. This means that, all effects considered, consumption and entry are positively related to the stock of physical capital and negatively related to the number of firms in the transition to the steady state.

3.2 The phase diagram in (K, n)

The fact that the system is four-dimensional, and the stable manifold is two-dimensional, poses some challenges to a qualitative understanding of the dynamics of transition. However, we can understand the dynamics of transition intuitively by concentrating on the projection onto the state-space (K, n), although there are insights to be gained by looking at the more familiar "Ramsey" projection (C, K).

We denote by E_1^s and E_2^s the lines in the four-dimensional space (C, e, K, n), which have the slope given by P_1^s and P_2^s, respectively, and pass through the equilibrium point. The one-dimensional lines E_1^s and E_2^s span the space E^s, which is tangent to the stable manifold at the steady-state equilibrium. To determine $E_1^s(E_2^s)$ we set $w_2^s = 0(w_1^s = 0)$, which implies that the dynamics is solely driven by $\lambda_1^s(\lambda_2^s)$. Hence,

$$E_1^s := \left\{(C, e, K, n) : \frac{C - C^*}{n - n^*} = \frac{\lambda_1^u - Cr_n}{\sigma l_1 - Cr_K}, \frac{e}{n - n^*} = \lambda_1^s, \frac{K - K^*}{n - n^*} = \frac{Cr_n}{\sigma l_1 - Cr_K}\right\},$$

and

$$E_2^s := \left\{(C, e, K, n) : \frac{C - C^*}{n - n^*} = \frac{\lambda_2^u - Cr_n}{\sigma l_2 - Cr_K}, \frac{e}{n - n^*} = \lambda_2^s, \frac{K - K^*}{n - n^*} = \frac{Cr_n}{\sigma l_2 - Cr_K}\right\},$$

evaluated in a neighborhood of the steady state. The dynamics of the system are driven by these two lines. When the system is close to the steady state, the equilibrium trajectories are asymptotically tanget to E_1^s and when further away the trajectories are parallel to E_2^s. This corresponds to the intuitive notion that the dynamics are driven at first more by the negative eigenvalue which is larger in absolute value, but since this dies away more quickly, the smaller eigenvalue predominates as you approach the steady state.

[13] See proof of Proposition 2 in the Appendix.

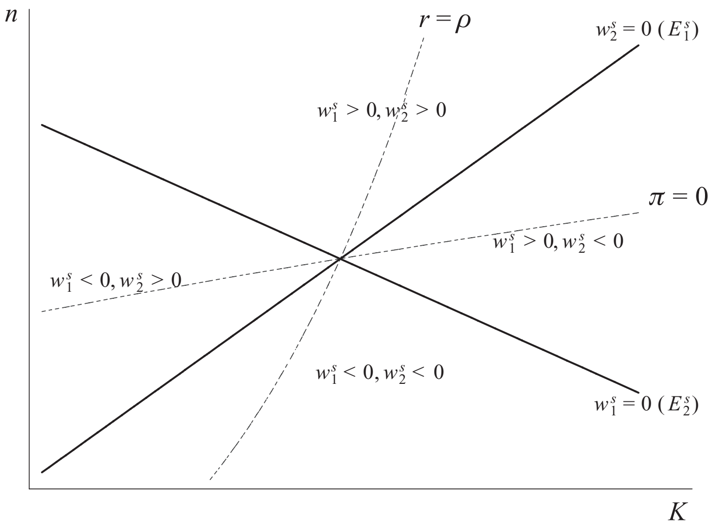

Figure 2 E_i^s lines in (K, n).

Proposition 2 *Qualitative characterization of the orbits belonging to the stable manifold. Consider an initial non-steady-state point* $(C(0), e(0), n(0), K(0))$*. The transition dynamics along the stable manifold will be as follows:*

(i) If the initial position of the two state variables lies on $E_i^s i = 1, 2,$ *then the control values jump to the corresponding values on* E_i^s *and the economy proceeds along* E_2^s *with maximum speed or* E_1^s *with minimum speed to the steady state.*
(ii) If the initial position of the two state variables does not lie on either E_i^s*, then the economy will move in the direction of* E_1^s *intially parallel to* E_2^s*. It will approach the steady state asymptotically tangent to* E_1^s*.*

We can now describe the two projections of E_i^s, $i = 1, 2$, in (K, n) and how they relate to our two reference curves (see Figure 2).

Proposition 3 *The projections of* E_1^s *and* E_2^s *in* (K, n) :

(i) The projections have the opposite slopes:

$$\left.\frac{dn}{dK}\right|_{E_1^s} = \frac{\sigma l_1 - Cr_K}{Cr_n} > 0, \quad \left.\frac{dn}{dK}\right|_{E_2^s} = \frac{\sigma l_2 - Cr_K}{Cr_n} < 0.$$

(ii) In general for $\gamma \in (0, \infty)$

$$\left.\frac{dn}{dK}\right|_{E_2^s} < 0 < \left.\frac{dn}{dK}\right|_{\pi=0} < \left.\frac{dn}{dK}\right|_{E_1^s} < \left.\frac{dn}{dK}\right|_{r=\rho}.$$

(iii) If $\gamma \to 0$ *then* $\left.\frac{dn}{dK}\right|_{E_2^s} \to -\infty$; *if* $\gamma \to \infty$ *then* $\left.\frac{dn}{dK}\right|_{E_1^s} \to \left.\frac{dn}{dK}\right|_{r=\rho}$.

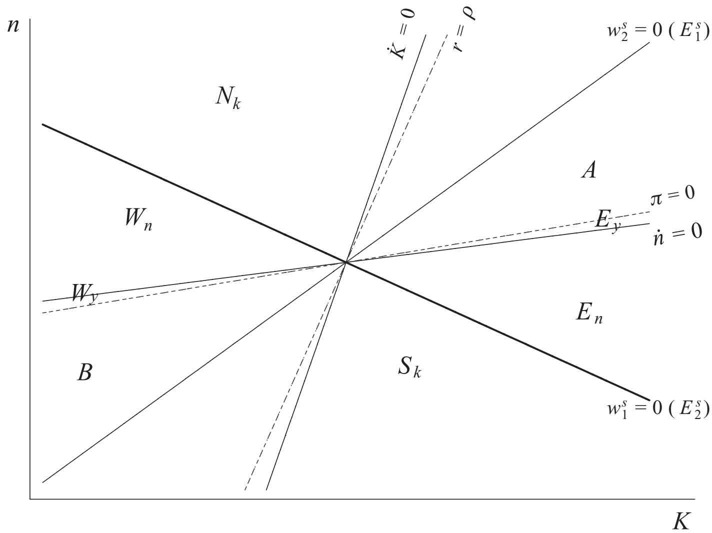

Figure 3 The projection of the isoclines for $\dot{K} = 0$ and $\dot{n} = 0$.

That is, in general the projection of E_1^s lies in between the two curves ($r = \rho$, $\pi = 0$), and the projection of E_2^s is negatively sloped. For $\gamma = 0$ (instantaneous free entry)[14] E_2^s is vertical and E_1^s corresponds with the zero profit line. For $\gamma = \infty$ (fixed number of firms), E_1^s is coincident with the $r = \rho$ curve and E_2^s is horizontal.

In general, when $\gamma \in (0, \infty)$, to describe the dynamics in full, we need to introduce two new lines: the isoclines $\dot{K} = 0$ and $\dot{n} = 0$ linearized around the steady state. In general, these isoclines will depend upon (C, e) and will not be invariant in (K, n). However, close to the steady state these depend only on (K, n) and, hence, can be used to characterize the equilibrium trajectories of the linearized system. Note that along the $\dot{n} = 0$ line, $q = 0$. This is because $\dot{n} = e = 0$ *if and only if* $q = 0$ (the net present value of incumbency is zero).

Proposition 4 *The tangents to the isoclines $\dot{K} = 0$ and $\dot{n} = 0$ passing through the steady-state are*

$$\left.\frac{dn}{dK}\right|_{\dot{n}=0} = \frac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n} > 0, \qquad \left.\frac{dn}{dK}\right|_{\dot{K}=0} = \frac{\sigma\lambda_1^s\lambda_2^s - Cr_K}{Cr_K} > 0.$$

As depicted in Figure 3, the $\dot{K} = 0$ isocline is positive in slope, and steeper than the $r = \rho$ line: the reason for this is that entry (exit) affects capital accumulation. The further south from the steady state we are, the higher the flow of entry. From (20), higher entry means less is available for investment and consumption, leading to less investment. Hence, capital accumulation might stop even though the marginal product of capital is above ρ. The converse happens north of the steady state. The $\dot{n} = q = 0$ isocline has a positive slope

[14] See Devereux, Head, and Lapham (1996), Heijdra (1988) and Coto-Martinez and Dixon (2003).

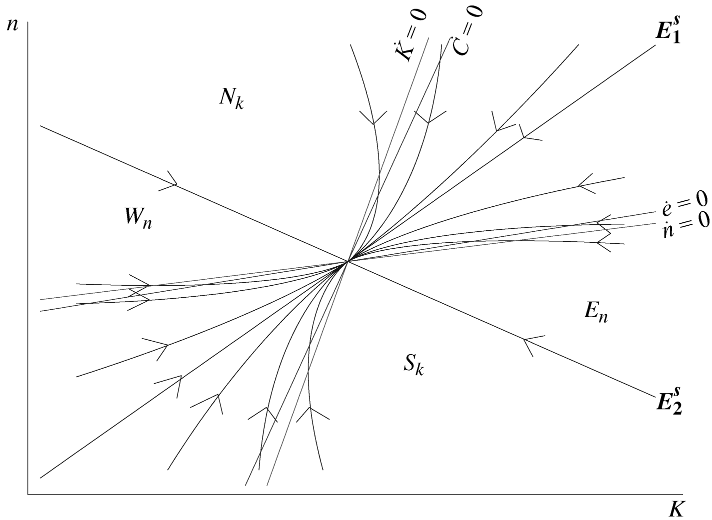

Figure 4 Representative trajectories for $0 < \gamma < \infty$.

that is less than the zero-profit $\pi = 0$ line because the *NPV* depends on the whole path of trajectories, not just the current flow of profit.

3.3 Dynamics in (K, n)

Now we have the isoclines, we can divide up the (K, n) projection of the stable manifold into regions depending on the types of trajectories, as depicted in Figure 4.[15] We can see that any trajectory that cuts the $\dot{K} = 0$ isocline must be vertical: any trajectory that cuts the $\dot{n} = 0$ must be horizontal. Any trajectory to the left of the $\dot{K} = 0$ must have capital increasing; any to the right decreasing. Any trajectory above the $\dot{n} = 0$ must have the number of firms decreasing and any below must have the number increasing.

If we start in a region between the isoclines, then we have monotonic dynamics in both variables: in region A we are to the right of the $\dot{K} = 0$ and above $\dot{n} = 0$, so that both variables are declining; in region B, we are to the left of the $\dot{K} = 0$ and below $\dot{n} = 0$, so that both variables are increasing. As the trajectories get close to the steady state, their slope will converge to that of E_1^s. We can treat the isoclines themselves as part of A and B, because trajectories starting from the isoclines also share the monotonicity.

In the regions outside A or B, we will in general get non-monotonic behavior in one variable: because the eigenvalues are real, there cannot be non-monotonic behavior in both variables along the same trajectory. If we are outside A and B, the only instance in which both variables are monotonic is when the initial position lies on E_2^s: in this case, the trajectory travels straight down E_2^s to the steady state, and the two state variables move in different directions (depending on whether they start above or below the steady state).

15 Figures 1–5 were generated using Maple for a specific calibration of the model (maplesoft.com). Details are available from the authors on request.

These two cases correspond to trajectories in which only the larger negative eigenvalue λ_2^s is active.

If we start from the regions strictly between $\dot{K} = 0$ and E_2^s, we will observe non-monotonic behavior in K. North-west of the steady state we have the region N_k. There are initially too many firms, $n > n^*$, and capital may be above or below the steady state. However, even if capital is above the steady state, the large number of firms boosts the marginal product of capital and encourages capital accumulation. This continues until the trajectory hits the $\dot{K} = 0$ isocline, and thenceforth enters region A and both variables decline towards the steady state. In the region S_k, south/south-west, the same occurs, but we have too few firms: capital will initially fall because the marginal product is low, until the $\dot{K} = 0$ is reached and the trajectory enters region B and both state variables increase to their steady state. Note that the regions N_k and S_k are both open sets: they do not include their boundaries E_2^s and $\dot{K} = 0$.

If we start from the region between E_2^s and $q = \dot{n} = 0$, we will observe non-monotonic behavior in entry and, hence, n. To the west of steady-state W_n, there is too little capital. This means that firms have negative NPV (we are above the $q = 0$ line), so that there will be exit until the $q = 0$ line is reached and then both variables enter region B and increase to the steady state. To the east of steady-state E_n, there is too much capital: this boosts firms NPV and induces entry, until the $q = 0$ line is met and the trajectory enters region A and both variables decline to the steady state.

Because the steady state is almost always approached along E_1^s, almost all trajectories must either approach through region A or B where both are increasing/decreasing. The only exception is where the initial position happens to lie on E_2^s. Hence, if the initial position lies outside $A \cup B$, the trajectory will move towards $A \cup B$ with one variable decreasing and one increasing: once it enters $A \cup B$ the dynamics become monotonic and there is a positive correlation between the state variables around the steady state.

We can now see why the $q = 0$ line is flatter than the $\pi = 0$. When the trajectory reaches the $q = 0$ line on the edge of A and there is too much capital, $NPV = 0$, despite $\pi > 0$, because the trajectory will cross the $\pi = 0$ line. Subsequently, strictly negative profits will be earned until the steady state is reached. Along the $q = 0$ line, the profits prior to reaching $\pi = 0$ are exactly offset by the subsequent losses. The opposite holds true when there is too little capital: the $q = 0$ line on the border of B is reached even though $\pi < 0$. This is because the trajectory will cross the $\pi = 0$ line and, subsequently, profits will be earned. Intuitively, we can think of π as representing "short-run" profits and q as "long-run" profits, and as we cross the $\pi = 0$ and $q = 0$ lines, the correlation between them changes.

We can now formally summarize the above insights:

Proposition 5 *Monotonous and non-monotonous transitional dynamics: Consider the following two subsets:*

$$N_k \equiv \left\{ (K, n) :\ n > n^* \text{ and } \frac{Cr_n}{\sigma l_2 - Cr_K} < \frac{K - K^*}{n - n^*} < \frac{Cr_n}{\sigma \lambda_1^s \lambda_2^s - Cr_K} \right\}, \tag{35}$$

$$S_k \equiv \left\{(K, n) : n < n^* \text{ and } \frac{Cr_n}{\sigma l_2 - Cr_K} > \frac{K - K^*}{n - n^*} > \frac{Cr_n}{\sigma \lambda_1^s \lambda_2^s - Cr_K}\right\} \quad (36)$$

and

$$W_n \equiv \left\{(K, n) :\ K < K^* \text{ and } \frac{\pi_K}{\gamma \lambda_1^s \lambda_2^s - \pi_n} < \frac{n - n^*}{K - K^*} < \frac{\sigma l_2 - Cr_K}{Cr_n}\right\} \quad (37)$$

$$E_n \equiv \left\{(K, n) :\ K > K^* \text{ and } \frac{\pi_K}{\gamma \lambda_1^s \lambda_2^s - \pi_n} > \frac{n - n^*}{K - K^*} > \frac{\sigma l_2 - Cr_K}{Cr_n}\right\}. \quad (38)$$

(i) If $(K(0), n(0)) \in N_k \cup S_k$, then $K(t)$ will adjust non-monotonically: it will increase and then decrease if starting in N_k; it will decrease an then increase if starting in S_k;

(ii) If $(K(0), n(0)) \in W_n \cup E_n$, then $n(t)$ will adjust monotonically; if starting in W_n it will increase and then decrease; if starting in R_n it will increase and then decrease;

(iii) If $(K(0), n(0)) \notin N_k \cup S_k \cup W_n \cup E_n$ then both state variables have monotonic trajectories.

3.4 Dynamics in (C, K)

Although the most intuition is gained by projecting the four-dimensional phase space onto (K, n), it is also illuminating to take a look at the conventional "Ramsey" projection onto (C, K).

Proposition 6 *Projections of E_i^s onto (C, K):*

$$\left.\frac{dC}{dK}\right|_{E_2^s} = \lambda_2^u > \lambda_1^u = \left.\frac{dC}{dK}\right|_{E_1^s} > 0.$$

In Figure 5 we depict the paths of two non-monotonic trajectories from Figure 4 in terms of (C, K) space.[16] For $\gamma \in (0, \infty)$, both E_i^s are upward sloping, so that consumption and capital move together: hence, when capital is non-monotonic, consumption will also be non-monotonic. Let us take the case where the initial capital stock is at its steady-state value, but the number of firms is well above the steady state. This will induce the capital stock to increase initially, and then decrease. Consumption will initially jump to a level below the steady state: this reflects the fact that it has a level of total assets below steady state (in terms of capital and firms the initial value is $V(0) = qn(0) + K(0)$ with $q < 0$) and the household wants to reduce the number of firms that it has. However, there is initially an increase in capital: the large number of firms causes the marginal product of capital to be high, which leads to capital accumulation. Consumption and capital move together with slope E_2^s. Eventually capital accumulation stops: this is when the trajectory in (K, n) reaches the $\dot{K} = 0$ isocline. However, consumption continues to increase even while capital

[16] These are the trajectories originating in N_K and S_K in Figure 4.

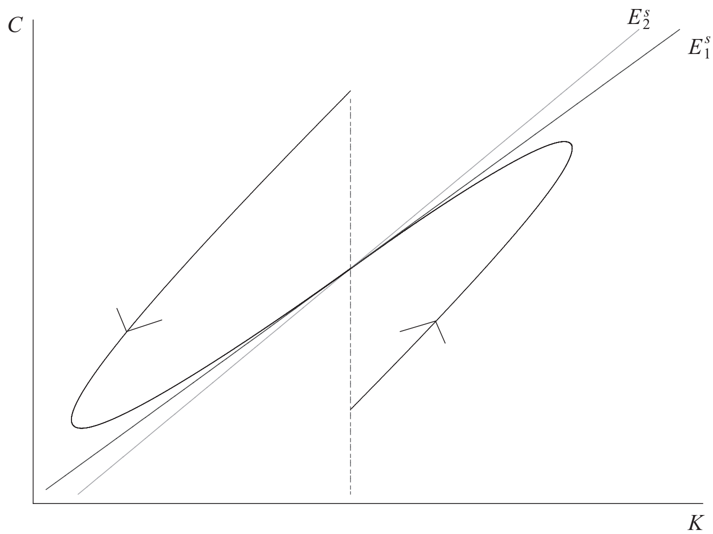

Figure 5 Trajectories where K is non-monotonic in (C, K)-space.

is falling (this represents the part of the trajectory in (K, n) between the $\dot{K} = 0$ isocline and the $r = \rho$ line). Eventually, capital and consumption both fall together along a path tangent to E_1^s. The non-monotonic trajectories with too many firms will all share this pattern: consumption starts low and initially increases with capital; there is a period when consumption continues to increase while capital falls; finally, both consumption and capital fall back down to the steady state. In the case where there are too few firms, the opposite occurs: because $q > 0$, consumption jumps above the steady state. Initially it declines with capital tangentially to E_2^s; then capital turns around and starts to increase; for a period consumption continues to decline, until both approach the steady state from below along E_1^s.

3.5 Firm size dynamics

Firm size dynamics along the stable manifold can be derived from the firm size equation (16), if we substitute the dynamics for the aggregate capital stock and the number of firms (32). In the neighborhood of the stationary equilibrium, the local dynamics for the firm size is given as

$$y(t) - y^* = \frac{1}{1-\nu}(\pi_K(K(t) - K^*) + \pi_n(n(t) - n^*)).$$

Therefore, the loci in the diagram (K, n) such that the size of firms is invariant (i.e. $y(t) = y^*$) is given by

$$\left.\frac{dn}{dK}\right|_{\dot{y}=0} = -\frac{\pi_K}{\pi_n},$$

which is coincident with the the zero-profit line $\pi(K, n) = 0$ (see Figure 1). Above that line we will have $\pi < 0$ and $y(t) < y^*$, and below the line $\pi > 0$ and $y(t) > y^*$. Therefore, the current flow of profits perfectly captures the size of the firm.

Using our previous analysis on the dynamics for the aggregate variables K and n we can also describe in which situations the dynamics of firm size can be non-monotonic. First, we define some extra subsets on the space (K, n):

$$W_y \equiv \left\{ (K, n): \; K < K^* \text{ and } -\frac{\pi_K}{\pi_n} < \frac{n - n^*}{K - K^*} < \frac{\pi_K}{\gamma \lambda_1^s \lambda_2^s - \pi_n} \right\}, \tag{39}$$

$$E_y \equiv \left\{ (K, n): \; K > K^* \text{ and } -\frac{\pi_K}{\pi_n} > \frac{n - n^*}{K - K^*} > \frac{\pi_K}{\gamma \lambda_1^s \lambda_2^s - \pi_n} \right\}. \tag{40}$$

Proposition 7 *Monotonous and non-monotonous transitional dynamics for the firm size:*

(i) If $(K(0), n(0)) \in W_n \cup W_y \cup E_n \cup E_y$ then $y(t)$ will adjust non-monotonically: it will increase and then decrease if starting in $W_n \cup W_y$; it will decrease and then increase if starting in $E_n \cup E_y$.

(ii) If $(K(0), n(0)) \notin W_n \cup W_y \cup E_n \cup E_y$ then $y(t)$ will adjust monotonically.

Note that for $0 < \gamma < \infty$ the non-monotonous adjustment of n is a sufficient (but not necessary) condition for the non-monotonous adjustment of firm size. If (K, n) belongs to $W_y(E_y)$ the transition path will cross the line $\pi = 0$ in its way to being tangent to E_1^s. This means that, although all the other variables vary monotonically, the variation in the size of firms shifts direction from being (decreasing) increasing to being decreasing (increasing). The dynamics for the size of firms will be similar when (K, n) belongs to $W_n(E_n)$. However, in this case, the number of firms will also adjust non-monotonically.

4 Conclusion

In this paper, we extend the Ramsey model and generalize existing approaches to entry by allowing for an explicit and fully transparent treatment of both the number and output per firm at the micro level along with the behavior of aggregate output, consumption and investment. We do this by allowing for an explicit treatment of two state variables: capital and number of firms. This contrasts with existing models that try to keep the number of state variables to one: either by allowing for the number of firms to vary but with no capital, having instantaneous free entry, or having a fixed number of firms. The reward for this additional complexity is that we have richer dynamic behavior. In particular, we have non-monotonic behavior in either one of the state variables (but not both because we have only real eigenvalues) resulting from the interaction of the state variables on each others' marginal product.

We believe that entry and exit have long been the Cinderella of dynamic general equilibrium analysis. This has largely been due to the technical difficulty of making the number of firms endogenous in a non-trivial way. We show that by adopting the dynamic entry model of Datta and Dixon (2002) it is possible to develop an intuitive and tractable dynamic general equilibrium model with the two state variables. Furthermore, although the model is inherently a four-dimensional system, the model can be represented graphically in two dimensions by projecting it onto the two-dimensional subspace of the state variables.

There are several ways to further develop the model in this paper. Most obviously, we can allow for imperfect competition, so that the long-run equilibrium is no longer optimal: in the steady state the zero profit condition will imply that average cost is greater than marginal cost, so that we have the standard Chamberlin–Robinson excess capacity result.

Appendix I

For more explicit proofs and related results, see Brito and Dixon (2007).

Proof of Proposition 1

The characteristic polynomial of the Jacobian matrix is $c(\lambda) = \lambda^4 - 2\rho\lambda^3 + M_2\lambda^2 - \rho(M_2 - \rho^2)\lambda + M_4$, where M_2 and M_4 are the sum of the principal minors of order 2 and 4 of the Jacobian J, where $M_2 = \rho^2 + \frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma}$ and $M_4 = \det(J) = \frac{C}{\sigma\gamma}(r_K\pi_n - r_n\pi_K)$. The characteristic polynomial can be equivalently written as $c(\lambda) = (\frac{\rho}{2})^4(w^2 + a_1 w + a_0)$, where

$$w \equiv \left(\lambda - \frac{\rho}{2}\right)^2 \left(\frac{\rho}{2}\right)^{-2}, \tag{41}$$

and $a_1 \equiv (\frac{\rho}{2})^{-2}(M_2 - \rho^2) - 2$ and $a_0 \equiv -a_1 + (\frac{\rho}{2})^{-4} M_4 - 1$. Then $c(\lambda) = 0$ if and only if $w^2 + a_1 w + a_0 = 0$. The roots of this polynomial on w are $w_{1,2} = -\frac{a_1}{2} \mp ((\frac{a_1}{2})^2 - a_0)^{\frac{1}{2}}$. If we substitute the expressions for the coefficients a_1 and a_0 we get

$$\left(\frac{\rho}{2}\right) w_{1,2} = \left(\frac{\rho}{2}\right)^2 - \frac{1}{2}\left(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma}\right) \mp \frac{1}{2}\left[\left(\frac{Cr_K}{\sigma} - \frac{\pi_n}{\gamma}\right)^2 + 4\frac{Cr_n}{\sigma}\frac{\pi_K}{\gamma}\right]^{\frac{1}{2}}.$$

Then, solving (41) for λ we obtain the eigenvalues as $(\lambda - \frac{\rho}{2})^{s,u}_{1,2} = \mp[(\frac{\rho}{2})w_{1,2}]^{\frac{1}{2}}$, which is equivalent to (28) and (29).

Next we demonstrate that the eigenvalues are real and satisfy (30). Recall that $r_K < 0$, $r_n > 0$, $\pi_n < 0$ and $\pi_K > 0$. Then $\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma} < 0$. The determinant of the Jacobian J is positive, as $\det(J) = \frac{C}{\sigma\gamma}(\pi_n r_K - \pi_K r_n) = \frac{C}{\sigma\gamma}(\frac{A}{n})^2(\frac{F_{KK}F_{LL} - F_{KL}^2}{n^2}) > 0$. Additionally, the discriminant is positive, as $\Delta = (\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma})^2 - 4\det(J) = (\frac{Cr_K}{\sigma} - \frac{\pi_n}{\gamma})^2 + 4\frac{C}{\sigma\gamma}\pi_K r_n > 0$, which implies that $\Delta^{\frac{1}{2}}$ is real and positive. It also implies that $\frac{1}{2}(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma}) + \frac{1}{2}\Delta^{\frac{1}{2}} < 0$ and that $(\frac{\rho}{2})^2 - \frac{1}{2}(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma}) - \frac{1}{2}\Delta^{\frac{1}{2}} > (\frac{\rho}{2})^2 > 0$. Then $[(\frac{\rho}{2})^2 - \frac{1}{2}(\frac{Cr_K}{\sigma} + \frac{\pi_n}{\gamma}) - \frac{1}{2}\Delta^{\frac{1}{2}}]^{\frac{1}{2}}$ is positive and real and larger than $\frac{\rho}{2}$. Therefore, the eigenvalues, given in (28) and (29), are real and verify (30). □

Proof of Proposition 2

Let $x \equiv (C(t) - C^*, e(t), K(t) - K^*, n(t) - n^*)^T$. Then, as $\det(J) \neq 0$, from the Hartman–Grobman theorem (Guckenheimer and Holmes 1990, p. 13), the local dynamics is topologically equivalent to the solution of the linear system $\dot{x}(t) = Jx(t)$. The eigenvalues associated to J have already been determined. Let $P = [P_2^u\, P_1^u\, P_1^s\, P_2^s]$ be the associated eigenvector matrix. Consider the vector $q = (q_2^u, q_1^u, q_1^s, q_2^2)$ such that $x = Pq$. As is well known, if we take the time derivatives of q and substitute $\dot{x}$ we get the product system $\dot{q}(t) = \Lambda q(t)$, where $\Lambda = P^{-1}JP$ is the Jordan matrix of J. Therefore, $\Lambda = \text{diag}(\lambda_2^u, \lambda_1^u, \lambda_1^s, \lambda_2^s)$. This system has the solution $q_i^u(t) = w_i^u e^{\lambda_i^u t}$, $q_i^s(t) = w_i^s e^{\lambda_i^s t}$ for $i = 1, 2$, where w_i^s and w_i^u are arbitrary constants. There is conditional convergence towards the steady state, if we set $w_1^u = w_2^u = 0$. First note that the eigenvectors associated with matrix J have the generic form

$$P_i^j = \left[\frac{\left(\rho - \lambda_i^j\right)\frac{Cr_n}{\sigma}}{l_i - \frac{Cr_K}{\sigma}}, \lambda_i^j, \frac{\frac{Cr_n}{\sigma}}{l_i - \frac{Cr_K}{\sigma}}, 1 \right]^T, \quad j = s, u, \quad i = 1, 2, \tag{42}$$

where $l_1 - \frac{Cr_K}{\sigma} > 0$ and $l_2 - \frac{Cr_K}{\sigma} < 0$, for any value of the parameters. Furthermore, the eigenvectors associated with the negative eigenvalues λ_1^s and λ_2^s are

$$P_1^s = \left[\frac{\lambda_1^u Cr_n}{\sigma l_1 - Cr_K}, \lambda_1^s, \frac{Cr_n}{\sigma l_1 - Cr_K}, 1 \right]^T, \quad P_2^s = \left[\frac{\lambda_2^u Cr_n}{\sigma l_2 - Cr_K}, \lambda_2^s, \frac{Cr_n}{\sigma l_2 - Cr_K}, 1 \right]^T.$$

Hence, the orbits, belonging to the space tangent to the stable manifold, are determined from:

$$\begin{pmatrix} C(t) - C^* \\ e(t) - 0 \\ K(t) - K^* \\ n(t) - n^* \end{pmatrix} = P_1^s w_1^s e^{\lambda_1^s t} + P_2^s w_2^s e^{\lambda_2^s t}, \tag{43}$$

where

$$w_1^s = \frac{l_1 - \frac{Cr_K}{\sigma}}{l_2 - l_1} \left[\left(\frac{l_2 - \frac{Cr_K}{\sigma}}{\frac{Cr_n}{\sigma}} \right)(K(0) - K^*) - (n(0) - n^*) \right], \tag{44}$$

$$w_2^s = \frac{l_2 - \frac{Cr_K}{\sigma}}{l_2 - l_1} \left[- \left(\frac{l_1 - \frac{Cr_K}{\sigma}}{\frac{Cr_n}{\sigma}} \right)(K(0) - K^*) + (n(0) - n^*) \right]. \tag{45}$$

If we assume that we know $K(t) - K^*$ and $n(t) - n^*$ at time t, then we obtain

$$w_1^s e^{\lambda_1^s t} = \frac{K(t) - K^* - P_2^s(3)(n(t) - n^*)}{P_1^s(3) - P_2^s(3)} \tag{46}$$

$$w_2^s e^{\lambda_2^s t} = \frac{-(K(t) - K^*) + P_1^s(3)(n(t) - n^*)}{P_1^s(3) - P_2^s(3)}. \tag{47}$$

As we usually know the initial values of the state variables at time $t = 0$, then we get (44) and (45) if we substitute the appropriate elements of P_1^s and P_2^s. If we substitute (43) for the non-predetermined variables C and e the expressions for $w_1^s e^{\lambda_1^s t}$ and $w_2^s e^{\lambda_2^s t}$ given by (46) and (47), we get a two-dimensional linear manifold that expresses the functional dependence with the state variables, along the stable manifold, E^s, as stated in the main text (33, 34). Part (i) follows from the definitions of E_i^s (only one eigenvalue is operative). Part (ii) follows from Hirsch and Smale (1974, p. 93). □

Proof of Proposition 3

From Proposition 2, we can use the definition of the schedules $w_1^s = 0$ and $w_2^s = 0$ to characterize the dynamics in the space (C, e, K, n) relative to the lines E_1^s and E_2^s. From Proposition 2, it follows that:

$$\left.\frac{dn^*}{dK}\right|_{E_1^s} = \frac{1}{P_1^s(3)} = \frac{\sigma l_1 - Cr_K}{Cr_n} > 0, \quad \left.\frac{dn}{dK}\right|_{E_2^s} = \frac{1}{P_2^s(3)} = \frac{\sigma l_2 - Cr_K}{Cr_n} < 0,$$

as E_1^s is associated with $w_2^s = 0$ and E_2^s with $w_1^s = 0$. The limiting cases follow from the inequalities taking the limits indicated. □

Proof of Proposition 4

If we consider (4) and (20) we see that the isoclines $\dot{K} = 0$ and $\dot{n} = 0$ depend on the control variables C and e, which depend on the K and n and, therefore, their projections in (K, n) are always shifting. However, we can determine loci in the space (K, n) that are analogous to the isoclines in a two-dimensional model: we can determine the loci where the trajectory belonging to the approximation to the stable manifold change direction; that is, the slope of $(n - n*)/(K - K^*)$ such that $d(K(t) - K^*)/dt = 0$ and the slope of $(n - n^*)/(K - K^*)$ such that $d(n(t) - n^*)/dt = 0$. For a full proof see Brito and Dixon (2007). □

Proof of Proposition 5

(i) Consider the solution for n in (43), and its derivative $\frac{d(n(t)-n^*)}{dt} = \lambda_1^s w_1^s e^{\lambda_1^s t} + \lambda_2^s w_2^s e^{\lambda_2^s t}$, where $e^{\lambda_1^s t} \geq 0$ and $e^{\lambda_2^s t} \geq 0$ for any $t \in \mathbb{R}_+$. If $\text{sign}(w_1^s) = \text{sign}(w_2^s)$ then $n(t)$ will converge monotonically towards the steady state n^*. If $\text{sign}(w_1^s) \neq \text{sign}(w_2^s)$ then $n(t) - n^*$ might change sign along the transition and, therefore, converge non-monotonically under certain conditions involving the parameters and the initial data.

Let us assume that

$$\begin{cases} \lambda_1^s w_1^s + \lambda_2^s w_2^s > 0 & \text{and} \quad w_2^s < 0, w_1^s > 0 \\ \lambda_1^s w_1^s + \lambda_2^s w_2^s < 0 & \text{and} \quad w_2^s > 0, w_1^s < 0 \end{cases}.$$

or, equivalently, that

$$\begin{cases} \dfrac{n(0) - n^*}{K(0) - K^*} < \dfrac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n} & \text{and} \quad w_2^s < 0, w_1^s > 0 \\ \dfrac{n(0) - n^*}{K(0) - K^*} > \dfrac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n} & \text{and} \quad w_2^s > 0, w_1^s < 0. \end{cases} \tag{48}$$

If conditions in (48) do not hold, then $\frac{d}{dt}(n(t) - n^*) \neq 0$ for any $0 \leq t < \infty$. If any of the two conditions in (48) hold then $\frac{\lambda_1^s w_1^s}{\lambda_2^s w_2^s} \geq -1$, which implies that there is a critical time

$$t_n = \frac{1}{\lambda_2^s - \lambda_1^s} \ln\left(-\frac{\lambda_1^s w_1^s}{\lambda_2^s w_2^s}\right) \geq 0,$$

such that $\frac{d}{dt}(n(t) - n^*)\big|_{t=t_n} = 0$, and the state variables verify

$$\begin{aligned} n(t_n) - n^* &= \left(\frac{\lambda_2^s - \lambda_1^s}{\lambda_2^s}\right) w_1^s e^{\lambda_1^s t_n} \\ K(t_n) - K^* &= \left(\frac{\gamma\lambda_1^s\lambda_2^s - \pi_n}{\pi_K}\right)\left(\frac{\lambda_2^s - \lambda_1^s}{\lambda_2^s}\right) w_1^s e^{\lambda_1^s t_n}, \end{aligned}$$

if we substitute $e^{(\lambda_2^s - \lambda_1^s)t} = -\frac{\lambda_1^s w_1^s}{\lambda_2^s w_2^s}$ into (44) and (45). In the (K, n)-space the locus where $\frac{d}{dt}(n(t) - n^*) = 0$,

$$\frac{n - n^*}{K - K^*} = \frac{n(t_n) - n^*}{K(t_n) - K^*} = \frac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n} > 0, \tag{49}$$

is coincident with the $\dot{n} = 0$ isocline defined in Proposition 4. Then there are two loci in the space (K, n), both delimited by line with slope $\frac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n}$ and by the E_1^s-projection, such that $n(t)$ will behave non-monotonically: W_n and E_n.

First, assume that $w_1^s < 0$, if $w_2^s > 0$ and that the time $t = 0$ value for the state variables, $(K(0), n(0))$, lies in subset W_n; that is, $\frac{n(0)-n^*}{K(0)-K^*} > \frac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n}$. Then the number of firms will fall along time until $t = t_n$, when its value crosses the $\frac{d}{dt}(n(t) - n^*) = 0$ line, and then increases afterwards. The path for n will next "cut," with a positive slope, curve $\pi(K, n) = 0$, and will tend asymptotically to the E_1^s-projection. If, instead, $w_1^s < 0$, $w_2^s > 0$ and the initial point does not lies in subset W_n, that is, if $\frac{n(0)-n^*}{K(0)-K^*} < \frac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n}$, then $(n(t) - n^*) < 0$, $\frac{d}{dt}(n(t) - n^*) > 0$ for any value of $0 \le t < \infty$ and $\lim_{t\to\infty} \frac{d}{dt}(n(t) - n^*) = 0$, and n will monotonically increase towards n^*.

Second, assume that $w_1^s > 0$, and $w_2^s < 0$, and that the initial values for the state variables belong to subset E_n; that is, $\frac{n(0)-n^*}{K(0)-K^*} < \frac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n}$. Then, the number of firms will increase until $t = t_n$, crossing the $\frac{d}{dt}(n(t) - n^*) = 0$ line, and decrease afterwards, crossing with a positive slope the curve $\pi(K, n) = 0$, and become asymptotically tangent to the E_1^s projection. If instead $w_1^s > 0$, and $w_2^s < 0$ but $\frac{n(0)-n^*}{K(0)-K^*} > \frac{\pi_K}{\gamma\lambda_1^s\lambda_2^s - \pi_n}$, then $(n(t) - n^*) > 0$, $\frac{d}{dt}(n(t) - n^*) > 0$ and for any value of $0 \le t < \infty$, $\lim_{t\to\infty} \frac{d}{dt}(n(t) - n^*) \to 0$. Hence n will monotonically decrease towards n^*.

(ii). Consider again (43) for K, and its derivative

$$\frac{d(K(t) - K^*)}{dt} = \frac{l_1 Cr_n}{\sigma l_1 - Cr_K} w_1^s e^{\lambda_1^s t} + \frac{l_2 Cr_n}{\sigma l_2 - Cr_K} w_2^s e^{\lambda_2^s t}.$$

As $e^{\lambda_1^s t} \ge 0$ and $e^{\lambda_2^s t} \ge 0$ for any $t \in \mathbb{R}_+$, and as $\frac{Cr_n}{\sigma l_1 - Cr_K} > 0$ and $\frac{Cr_n}{\sigma l_2 - Cr_K} < 0$, then $\text{sign}(w_1^s) = \text{sign}(w_2^s)$ and $K(t)$ may converge non-monotonically towards the steady state K^*. If instead $w_1^s > 0$ and $w_2^s < 0$($w_1^s < 0$ and $w_2^s > 0$) then $K(t)$ will decrease (increase) monotonically towards K^*. (This is what happens in quadrants W_n and E_n.) To determine the conditions and the loci in the (K, n)-space for which we may have $\frac{d}{dt}(K(t) - K^*) = 0$, we follow the same procedure as previously. Let

$$\begin{cases} \dfrac{n(0) - n^*}{K(0) - K^*} < \dfrac{\sigma\lambda_1^s\lambda_2^s - Cr_K}{Cr_n} & \text{and} \quad w_1^s < 0, w_2^s < 0 \\ \dfrac{n(0) - n^*}{K(0) - K^*} > \dfrac{\sigma\lambda_1^s\lambda_2^s - Cr_K}{Cr_n} & \text{and} \quad w_1^s > 0, w_2^s > 0, \end{cases} \tag{50}$$

then there is a critical time

$$t_K = \frac{1}{\lambda_2^s - \lambda_1^s} \ln\left(-\frac{\lambda_1^s w_1^s(\sigma l_2 - Cr_K)}{\lambda_2^s w_2^s(\sigma l_1 - Cr_K)}\right) \ge 0,$$

such that $\frac{d}{dt}(K(t) - K^*)|_{t=t_K} = 0$, and substituting the critical time in (44) and (45) we obtain the set of values for (K, n) such that this condition holds:

$$\frac{n - n^*}{K - K^*} = \frac{n(t_K) - n^*}{K(t_K) - K^*} = \frac{\sigma\lambda_1^s\lambda_2^s - Cr_K}{Cr_n}. \tag{51}$$

Again, we can find two loci in the space (K, n), both delimited by a line with slope $\frac{\sigma\lambda_1^s\lambda_2^s - Cr_K}{Cr_n}$ and the E_1^s projection, such that $K(t)$ will behave non-monotonically: N_k or S_k.

If the initial and steady-state values for the state variables $(K(0), n(0))$ and (K^*, n^*) satisfy the conditions given by (50) then the saddle path will cut the line given by (51) at time t_K, changing the direction of evolution of the K variable. After a while it will cross line $r(K, n) = \rho$ and converge asymptotically to (K^*, n^*). If conditions given by (50) do not hold, then $\frac{d}{dt}(K(t) - K^*) \neq 0$ for any $0 \le t < \infty$ and the adjustment of K will be monotonic. Those conditions hold in the areas N_k and S_k given analytically in (35) and (36): in the first case K will decrease until it reaches the $\dot{K} = 0$ line and increases afterwards, and in the second case it has the opposite evolution. □

Proof of Proposition 6

If $w_1^s = 0$, then from (44) and Proposition 2 the projections of E_2^s are:

$$\left.\frac{dC}{dK}\right|_{E_2^s} = \frac{P_2^s(1)}{P_2^s(3)}, \quad \left.\frac{de}{dn}\right|_{E_2^s} = \frac{P_2^s(2)}{P_2^s(4)},$$

yielding the values reported. Likewise for E_1^s. □

Proof of Proposition 7

Analogous to Proposition 5.

References

Aloi, M., and H. Dixon (2002), "Entry dynamics, capacity utilisation and productivity in a dynamic open economy," University of York Discussion Paper 02/05, University of York, UK.

Baumol, W. J., J. C. Panzar, and R. D. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, New York, London: Harcourt Brace Jovanovich.

Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2005), "Business cycles and firm dynamics," Society for Economic Dynamics. Available from: www.economicdynamics.org/society.htm.

Brito, P., and H. Dixon (2007), "Entry and the accumulation of capital: A two state-variable extension to the Ramsey model," Cardiff Business School Working Paper E2007/16, Cardiff.

Coto-Martinez, J., and H. Dixon (2003), "Profits, markups and entry: Fiscal policy in an open economy," *Journal of Economics Dynamics and Control* **27,** 573–97.

Das, S., and S. Das (1996), "Dynamics of entry and exit of firms in the presence of entry adjustment costs," *International Journal of Industrial Organization* **15,** 217–41.

Datta, B., and H. Dixon (2002), "Technological change, entry, and stock market dynamics: An analysis of transition in a monopolistic economy," *American Economic Review* **92,** 231–5.

Devereux, M. B., A. C. Head, and B. M. Lapham (1996), "Monopolistic competition, increasing returns, and the effects of government spending," *Journal of Money, Credit and Banking* **28,** 233–54.

Gort, M., and S. Klepper (1982), "Time paths in the diffusion of product innovation," *Economic Journal* **92,** 630–53.

Guckenheimer, J., and P. Holmes (1990), *Nonlinear Oscillations and Bifurcations of Vector Fields*, 2nd edn, New York: Springer-Verlag.

Heijdra, B. (1998), "Fiscal policy multipliers: The role of monopolistic competition, scale economies, and intertemporal substitution in labor supply," *International Economic Review* **39,** 659–96.

Hirsch, M., and S. Smale (1974), *Differential Equations, Dynamical Systems, and Linear Algebra*, London: Academic Press.

Hopenhayn, H. A. (1992), "Entry, exit and firm dynamics in long run equilibrium," *Econometrica* **60,** 1127–50.

Hornstein, A. (1993), "Monopolistic competition, increasing returns to scale, and the importance of productivity shocks," *Journal of Monetary Economics* **31,** 299–316.

Howrey, E. P., and R. E. Quandt (1968), "The dynamics of the number of firms in an industry," *Review of Economic Studies* **35,** 349–53.

McKenzie, L. W. (1959), "On the existence of general equilibrium for a competitive market," *Econometrica* **27,** 54–71.

Myers, M. G., and E. R. Weintraub (1971), "A dynamic model of firm entry," *Review of Economic Studies* **38,** 127–9.

Novshek, W., and H. Sonnenschein (1987), "General equilibrium with free entry: A synthetic approach of the theory of perfect competition," *Journal of Economic Literature* **25,** 1281–1306.

Okuguchi, K. (1972), "A dynamic model of firm Entry: Comment," *Review of Economic studies* **39,** 521–2.

Smith, V. (1974), "Optimal costly firm entry in general equilibrium," *Journal of Economic Theory* **9,** 397–417.