Supplementary materials for Nikolaos Vettas, "Investment dynamics in markets with endogenous demand," The Journal of Industrial Economics, vol. 48, no. 2, June 2000, pp. 189-203.

● *The topics in this Appendix are linked to footnotes in the published article. Numbered paragraphs here refer to the corresponding footnote numbers in the text.*

A4. Willingness to pay is increasing in past sales. Multiplicities of equilibria in related contexts, in particular due to network effects, have been discussed in the literature. If p_t were increasing in current sales (if we had $p_t = p(x_t)$ instead of $p_t = p(x_{t-1})$), the market could possibly coordinate in a number of different states. We could have equilibria with high demand and high entry, as well as low demand and low entry. While in some markets the level of current sales may also be important, x_{t-1} is used here as a proxy for the level of the consumers' overall exposure to the sales volume of the product.

A9. "Microfoundations" for the demand. An information-based description of consumer behavior is consistent with the assumed pattern. Suppose, for example, that consumers' willingness to pay is an increasing function of the expectation of θ, due to network effects. Further, consumers have access to the same information as firms and learn in a Bayesian fashion. Then their willingness to pay is an increasing function of past capacities, since expected θ is increasing in x_{t-1} as long as there is no overshooting. I thank a referee for pointing this out.

A10. Discussion: demand uncertainty and endogenous demand. Given that there are two key components in the model, demand uncertainty for the firms and dependence of demand on past sales, it is important to discuss the role of each in the analysis. First, suppose that there is only demand uncertainty but demand is fixed. This situation, which corresponds to Rob [1991], is admitted in the analysis as a special case ($m = 0$). In this case, the dynamics are driven exclusively by demand learning. The determining factor is the distribution function F. As long as its hazard rate is not decreasing, entry is decreasing over time (reflecting the fact that with more invested capacity there is a greater danger of overshooting). For example, with F uniform, both the competitive and the optimal cumulative path will be concave (with optimal entry always exceeding the competitive by a level reflecting the one externality that remains in this framework, learning the demand). When we also consider the second key component of the model, that demand may grow endogenously, we can characterize a second externality associated with cultivating the market and show that the expansion paths will be qualitatively different.

Second, assume that demand is an increasing function of past sales but firms face no demand uncertainty. In our analysis, demand uncertainty implies that capacity expansion is gradual: As new information about θ is revealed in every period, additional entry is

attracted. If firms faced no demand uncertainty, entry would be instantaneous and cover the entire market (this can be seen formally if F becomes a trivial distribution with all its mass on a given value) - clearly in this case there is no room for learning. Clearly, a study of the dynamic interrelation between entry and demand would not be possible in an instantaneous adjustment scenario. Note that if demand learning was not present in the model, adjustment costs or firm heterogeneity could also lead to gradual expansion.

A12. **Proof of Proposition 1**. (3) follows from $R_t(x_{t-1}, y_t) = c$. The probability of overshooting is $F(x_{t-1} + y_t \mid x_{t-1} < \theta)$ and gives zero profit. With the complementary probability, the firm has current period profit equal to $p(x_{t-1})$ and a continuation payoff which is equal to c, because profits are competed away and $R_t(x_{t-1}, y_t) = c$ applies *in the following period as well*. Existence and uniqueness can then be established using elementary methods. It can be shown that the RHS of (3) is strictly decreasing and continuous in y_t and becomes higher than c for $y_t = 0$ (using $p_1 > c(1 - \beta)$) and lower than c for y_t too high. For details see Vettas [1998] where, in turn, the proof follows the methodology developed in Rob [1991] for the constant demand case.

A16. **Proof of Proposition 2: The solution to the dynamic programming problem**. For simplicity, we suppress the time subscripts in these calculations and write x for x_{t-1} and y for y_t. Differentiating the RHS of (5) with respect to y and rearranging, we obtain the following first-order condition:

$$c = [p(x) + \beta V'(x+y)]\frac{1 - F(x+y)}{1 - F(x)} - \beta \left[V(x+y) - \frac{p(x+y)(x+y)}{1-\beta} \right] \frac{f(x+y)}{1 - F(x)}.$$

Next, using the "envelope" theorem we obtain:

$$V'(x) = [p'(x)(x+y) + p(x) + \beta V'(x+y)]\frac{1 - F(x+y)}{1 - F(x)} +$$

$$[p(x)(x+y) + \beta V(x+y)]\frac{[1 - F(x+y)]f(x) - [1 - F(x)]f(x+y)}{[1 - F(x)]^2} +$$

$$\int_x^{x+y} [p'(x)\frac{\theta f(\theta)}{1 - F(x)} + [p(x) + \frac{\beta}{1-\beta}p(\theta)]\frac{\theta f(\theta)f(x)}{[1 - F(x)]^2}]d\theta +$$

$$[p(x) + \frac{\beta}{1-\beta}p(x+y)]\frac{(x+y)f(x+y)}{1 - F(x)} - \left[\frac{p(x)x}{1-\beta}\frac{f(x)}{1 - F(x)} \right].$$

2

Using the first order condition to substitute for the term $\beta[V(x+y) - \frac{p(x+y)(x+y)}{1-\beta}] \frac{f(x+y)}{1-F(x)}$ in the envelope equation and manipulating the resulting equation we have:

$$V'(x) = c + p'(x)(x+y)\frac{1-F(x+y)}{1-F(x)} + \int_x^{x+y} p'(x)\frac{\theta f(\theta)}{1-F(x)}d\theta +$$

$$\frac{f(x)}{1-F(x)}\left\{[p(x)(x+y)+\beta V(x+y)]\frac{1-F(x+y)}{1-F(x)} - \frac{p(x)x}{1-\beta} + \int_x^{x+y}[p(x)+\frac{\beta}{1-\beta}p(\theta)]\frac{\theta f(\theta)}{[1-F(x)]}d\theta\right\}.$$

The next step is to use the value equation that follows from (5) to substitute for the term $[p(x)(x+y)+\beta V(x+y)]\frac{1-F(x+y)}{1-F(x)}$. We obtain:

$$V'(x) = c + p'(x)(x+y)\frac{1-F(x+y)}{1-F(x)} + \int_x^{x+y} p'(x)\frac{\theta f(\theta)}{1-F(x)}d\theta + \frac{f(x)}{1-F(x)}[V(x)+cy-\frac{p(x)x}{1-\beta}].$$

Taking this expression one period forward, substituting in the first order condition for the second term of its RHS, and rearranging terms we finally obtain the Euler equation (6).

A17. **Optimal expansion under linear demand growth.** Under (2), (6) becomes:

$$c = (p_1 + mx_{t-1} + \beta c)\frac{1-F(x_{t-1}+y_t)}{1-F(x_{t-1})} + \beta c y_{t+1}\frac{f(x_{t-1}+y_t)}{1-F(x_{t-1})} +$$

$$+\beta m\left[(x_{t-1}+y_t+y_{t+1})\frac{1-F(x_{t-1}+y_t+y_{t+1})}{1-F(x_{t-1})} + \int_{x_{t-1}+y_t}^{x_{t-1}+y_t+y_{t+1}}\frac{\theta f(\theta)}{1-F(x_{t-1})}d\theta\right]. \quad (A17.1)$$

Comparing optimal expansion paths with $m' > m$. (See Figure A2 in this Appendix for an illustration.) Formally, we have the following result: Let $m' > m$. The optimal capacity path, x_t^o, for m' is always "above" the path for m. That is, using the notation $x_t^o(m)$ to indicate the dependence of capacity on m, $x_t^o(m') > x_t^o(m)$ for $t = 1, 2, 3, ...$

The key for the proof is that a larger m increases expected profitability in the market without increasing the cost of overshooting. If overshooting occurs, the firm does not lose its flow profit. Its only cost is that is has invested in capacity that will not be used. Therefore, the cost of overshooting depends only on the cost of capital, c, and not on the demand parameter m. On the other hand, a higher m increases the incentive to expand for two reasons. First, the current price p_t (and, thus, the expected profit per unit of capacity invested) is higher. Second, a higher m yields a higher expected "return" in terms of increasing future demand for each unit of capacity that is currently invested.

A19. **The equation for the optimal investment path (under linear demand growth).** When F is uniform on $[\underline{\theta}, \bar{\theta}]$, from equation (A17.1) in this Appendix we obtain for $t = 2, 3, ...$

$$(\bar{\theta} - x_{t-1})c = (p_1 + mx_{t-1} + \beta c)(\bar{\theta} - x_t) + \beta(x_{t-1} - x_t)\left[\frac{m}{2}(x_{t-1} + x_t) + c\right] + \beta mx_{t+1}(\bar{\theta} - x_{t-1})$$
$$(A19.1)$$

which is a non-linear second order difference equation in x_t. Rearranging terms, (A19.1) can be written as

$$a_1 x_{t+1}^2 + a_2 x_{t+1} + a_3 = 0,$$

where $a_1 = \beta m/2$, $a_2 = -\beta(m\bar{\theta}+c)$, $a_3 = c(\bar{\theta}-x_{t-1}) - (p_1 + mx_{t-1} + \beta c)(\bar{\theta}-x_t) + \beta cx_t + \beta\frac{m}{2}x_t^2$. It is easy to check that the larger of the two roots of the above equation is always above $\bar{\theta}$. Such values clearly cannot belong to the optimal path. Therefore, only the smaller root is admissible and it gives (7) which has to be satisfied by the optimal investment path.

A21. **Equation (7) and the optimal solution.** (7) is a second-order equation and we have only one initial condition, x_0. There are some additional properties that the optimal path should satisfy. First, obviously $x_t \in [0, \bar{\theta}]$ for all t. Second, x_t is increasing. Third, the optimal capacity path is always above the corresponding competitive path, that is, it cannot be smaller than the RHS of the capacity equation in (4) - this important property is formally proved in Proposition 3. In addition, as shown above, competitive entry does not stop until overshooting occurs. Since the optimal path has to be above the competitive path, this also implies (for uniform F) that capacity tends over time to the upper bound of the domain. In what follows, this can be used as a second boundary condition. Summarizing, the optimal path will be analyzed using the information that it satisfies (7), monotonicity, $x_t \in [0, \bar{\theta}]$, and the two conditions $x_0 = 0$ and $x_t \to \bar{\theta}$ as $t \to \infty$.

A23. **The optimal solution and the dynamical system.** Substituting $x_{t-1} = x_1 = x_{t+1}$, we see that (7) has two fixed points, $\bar{\theta}$ and $(c - p_1 - \beta c)/m(1 + \beta)$. It is easy to see that the second fixed point is less than zero (using $p_1 > c(1 - \beta)$), and is not admissible since capacities satisfy $x_t \geq 0$. The next step is to study the dynamics of the system. First the local properties around $\bar{\theta}$ need to be analyzed since we know that the optimal solution satisfies $x_t \to \bar{\theta}$ as $t \to \infty$. Then the global properties of (7) should be examined to obtain some further information about the optimal path.

It is helpful to rewrite (7) as a system of first-order equations. Defining $z_t \equiv x_t$, $w \equiv x_{t-1}$, and $\varphi(x_{t-1}, x_t)$ as the RHS of (7), we obtain:

$$\begin{cases} w_{t+1} = z_t \\ z_{t+1} = \varphi(w_t, z_t) \end{cases}. \tag{A23.1}$$

First we examine the local stability of $\bar{\theta}$. The Jacobian, evaluated at $z_t = w_t = \bar{\theta}$, is

$$\begin{bmatrix} 0 & 1 \\ -1/\beta & (p_1 + (1+\beta)m\bar{\theta} + 2\beta c)/\beta c \end{bmatrix}.$$

By considering the characteristic equation and the fact that the determinant of the Jacobian is greater than 1 and its trace positive, we can show that the Jacobian has two eigenvalues, λ_1, λ_2, both real, with $\lambda_1 \in (0,1)$ and $\lambda_2 > 1$. It follows that, with respect to its local stability properties, $\bar{\theta}$ is a saddle. Further, the stable eigenvalue λ_1 can be calculated as

$$\lambda_1 = \frac{1}{2\beta c} \left[p_1 + (1+\beta)m\bar{\theta} + 2\beta c - \sqrt{(p_1 + (1+\beta)m\bar{\theta} + 2\beta c)^2 - 4\beta c^2} \right] \tag{A23.2}$$

and the corresponding eigenvector is:

$$\lambda_1 \cdot (1/\lambda_1, 1).$$

Since $\bar{\theta}$ is a saddle, *in the linearized system* there is a unique path, the "saddle path", converging to $\bar{\theta}$. Its equation is

$$(w_t - \bar{\theta}) = \frac{1}{\lambda_1}(z_t - \bar{\theta})$$

which, since we have defined $z_t = x_t$ and $w_t = x_{t-1}$, can be rewritten as $x_t = \lambda_1 x_{t-1} + \bar{\theta}(1 - \lambda_1)$. The saddle path is the stable (invariant) subspace for the linearized system (see e.g. Guckenheimer and Holmes [1983], p.17). Since $\lambda_1 \in (0,1)$, for $x_0 = 0$ we obtain:

$$x_t = \bar{\theta}[1 - (\lambda_1)^t]. \tag{A23.3}$$

Note that since $\bar{\theta}$ is hyperbolic (eigenvalues different from 1 in absolute value), the linearization is a good local approximation. Since $\bar{\theta}$ is a saddle we know that any solution close to $\bar{\theta}$ diverges unless it approaches from a single direction, that of the stable eigenvector. Thus, since the optimal path converges to $\bar{\theta}$ it has to follow locally this direction. (Note that, as discussed below, for $m = 0$ the system becomes linear and (A23.3) itself is the optimal path.) These considerations also explain why in the numerical examples the system is very sensitive: points that approach $\bar{\theta}$ from the wrong direction diverge.

The analysis so far has been only local. We now turn to the global properties of (A23.1). By using a phase diagram, that is, a diagram showing the direction of increase for the variables, we are able to identify a specific region where the optimal path necessarily belongs. We examine first the phaselines, that is, the points where one of the two variables is stationary. From (A23.1) it follows that the $w_t = w_{t+1}$ locus is the $w_t = z_t$ equation, that is, the $45°$ line of the (w_t, z_t) plane. The $z_t = z_{t+1}$ locus, on the other hand, is $z_t = \varphi(w_t, z_t)$ which can be solved explicitly for w_t:

$$w_t = \frac{\bar{\theta}(p - c + \beta c) - (p + \beta c - \beta m\bar{\theta})z_t - \beta m z_t^2}{m z_t - m\theta - c} \tag{A23.4}$$

Finally, we can complete the phase diagram by identifying whether w_t and z_t are increasing or decreasing above and below the phaselines. Clearly for points above (below) the $w_t = z_t$ line we have $w_{t+1} < w_t$ ($w_{t+1} > w_t$). It is also easy to establish that for points above (below) the $z_{t+1} = z_t$ line we have $z_{t+1} < z_t$ ($z_{t+1} > z_t$). We conclude that the phase diagram is as in Figure A4 in the Appendix, where we use arrows to indicate the direction of increase for each variable under (A23.1).

We can now use the above properties to characterize the optimal solution. First, since the optimal solution is increasing and satisfies $x_t \in [0, \bar{\theta}]$ for all t, it follows from the vector field that *the optimal path is restricted in the area (A)* defined by $z_t = \bar{\theta}$, $y_t = 0$ and the $z_t = z_{t+1}$ line. Points outside A clearly cannot belong to the optimal path. Defining A more precisely, we can show that the $z_{t+1} = z_t$ locus intersects the $w_t = 0$ axis only at one positive value, $\bar{z}$, which can be calculated as

$$\bar{z} = \frac{1}{2\beta m}\left[\sqrt{(p + \beta c + \beta m\bar{\theta})^2 - 4\beta m\bar{\theta}c} - (p + \beta c + \beta m\bar{\theta})\right]. \tag{A23.5}$$

We can further show that for $z_t \in [\bar{z}, \bar{\theta}]$ the $z_{t+1} = z_t$ locus is positive, increasing and convex. (Note that $z_t = z_{t+1}$ diverges to $+\infty$ as $z_t \to [(m\bar{\theta}+c)/m]^- > \bar{\theta}$ and that it intersects $w_t = 0$ at a second, negative point. Also the saddle path for the linearized system is, in area A, "between" the $z_t = z_{t+1}$ and the $z_t = \bar{\theta}$ equations.)

So far we have concluded that the optimal path cannot be outside of area A and should approach $\bar{\theta}$ from a specific direction, otherwise it diverges. Proceeding further, we can characterize the optimal solution as follows. Intuitively, it seems obvious that there is a unique path in A converging to $\bar{\theta}$. Points off this path diverge, as indicated by the arrows in the vector field. More precisely, the optimal solution consists of points that lie on what is referred to as the *stable manifold* of $\bar{\theta}$ (see e.g. Guckenheimer and Holmes [1983], p.18). The "stable manifold theorem" guarantees the existence and uniqueness of an one-dimensional stable manifold for the system under consideration. A key property is that points on the

6

(x_{t-1}, x_t) space *converge to $\bar{\theta}$ if and only if they belong to the stable manifold of $\bar{\theta}$.* This completes the analysis.

Note that while, as usual in this class of problems, this manifold cannot be generally expressed in closed form, we know that it satisfies the following properties: It is between the $z_t = z_{t+1}$ and $z_t = \bar{\theta}$ equations in area A and can be expressed as a smooth line $w_t = h(z_t)$. It is tangent to the stable subspace (saddle path) of the linearized system at the fixed point: $h(\bar{\theta}) = \bar{\theta}$ and $h'(\bar{\theta}) = 1/\lambda_1$. It is also invariant (all points on the manifold tend to the fixed point under forward iteration of the dynamical system) which implies that $z_t = h(\varphi(h(z_t), z_t))$.

A24. **Discussion of two cases.**

Digital television. A case of endogenously changing demand arises when network effects are present. A recent interesting example is provided by the transition to a digital television (DTV) standard. This transition involves costly decisions by TV manufacturers, broadcasters, and consumers - and it is important to note that some of these payers are "small," while others are "large" (e.g. RCA which owned NBC was a big player in the TV market). Whereas there are many aspects in this problem, such as a choice between competing standards (see the related analysis of Farrell and Shapiro [1992]), two key factors are demand uncertainty and externalities. These considerations underlie a central problem that the Federal Communications Commission had to face in its effort to make the new DTV market work smoothly: "One difficulty is the "chicken-and-egg" relationship between transmission and reception. Broadcasters are not eager to invest significant sums to broadcast a signal that no one can receive. Manufacturers are reluctant to build - and consumers will be reluctant to buy - receivers for which there is no programming." [from a Separate Statement of FCC commissioner Susan Ness, April 3, 1997, regarding MM Docket No. 87-268.] The FCC's 1997 decision to require adoption by TV stations faster than the stations might have chosen seems appropriate. In particular, the FCC adopted rules to guarantee that there will be three or four network-affiliated digital signals in the top ten markets in the U.S. by November 1, 1999 as well as specified other targets to monitor the development of the market (See FCC's 6th Report and Order on MM Docket No. 87-268 - FCC 97-115, adopted 4/3/97, released 4/21/97; "Advanced Television Systems and their Impact Upon the Existing Television Broadcast Service"). As explained by the then FCC chairman Reed E. Hundt, this plan "scuttles the laissez-faire approach of the 1992 decision. Now we rely on the lead dogs to move the transition, which gives the country the biggest bang with the smallest buck." (See Separate Statement, April 3, 1997). Hundt further explained that a focus in multiple TV signals in each market is critical since "consumers won't buy TV sets for a single improved signal" and expressed the feeling that an even stricter rule (for 18 months adoption) may be

preferable since, among other things, "it would have given manufacturers the certainty they need to build TV sets in massive amounts...".

While the model in this paper certainly does not capture all aspects of the transition to DTV, its implications offer some insights regarding such policy issues. In particular, the model characterizes the dual externalities, regarding learning and "cultivating" the demand, that lead to inefficiently low entry. When incentives are created for early entry, not only is uncertainty resolved more quickly but there is a dynamic feedback with demand increases attracting further entry. Naturally, there are now a number of growth forecasts for DTV (see e.g. Bayus [1993] for a discussion of such estimates). The central idea in this paper implies that demand forecasts cannot be independent of factors that affect firms' entry into the market, such as the FCC order discussed above (and, in fact, a recognition of demand endogeneity appears central in such decisions). Note that the DTV transition issue is also encountered in other countries. Depending on how strongly early entry will be encouraged in different markets, we may observe cumulative diffusion paths that tend to be either more S-shaped, when a rather "free market" approach is taken, or more "concave" when such incentives are stronger.

Soccer in the U.S. Recent examples of opening new sports markets, such as professional women's basketball leagues, soccer in Japan, or American football in Europe, highlight a demand endogeneity problem. In particular, the case of soccer in the U.S. is interesting, with consumers' interest being dependent on the involvement of television networks and sponsors, and vice versa. Further, as a recent article notes, "many companies don't want to get involved in soccer until they see the payoff, but they won't see the payoff unless more companies get involved." (*Wall Street Journal,* June 12, 1998, "Why soccer will (won't) catch on.") While there are no "policy" issues here in the sense of government intervention, it was recognized that a "laissez-faire" approach was not appropriate since the experiences needed for demand to increase could not be generated by individual firms. A *coordinated* effort, on the other hand, could result to a large demand increase. A number of steps were taken to help the growth of the market, starting with the decision to host the 1994 World Cup, a multi-million dollar event, in the U.S. An important part of this decision was the need for the market to be "cultivated" and the idea that "potential" demand was important rather than its current level. The start of a new professional league (MLS) followed, building on the interest generated by the Cup. In its first year, the league showed a $19 million loss but this loss was $3 million smaller than projected, demonstrating again that the organizers had recognized that demand could grow and profits increase only after consumers have enough experiences (*Wall Street Journal,* July 15, 1997, "New leagues go for central ownership."). At the same time sponsors became increasingly more interested: While before it had almost

no soccer business, "after the 1994 World Cup played to sellout crowds in the U.S., Nike launched its soccer assault. The company's world-wide soccer footwear and apparel sales are projected at more than \$425 million for fiscal 1998, up about 50% from a year earlier..." *Wall Street Journal,* October 22, 1997, "Nike kicks in millions to sponsor soccer in U.S." While demand uncertainty is still unresolved, the recognition of demand endogeneity and that, while initial demand was too low to justify entry by individual firms, a coordinated effort may succeed, are important from the point of view of the present analysis.

A25. **Proof of Proposition 3: The optimal path is "above" the competitive path**. We use function $R(x, y)$. Let $x_{t-1}^e \leq x_{t-1}^o < \theta$. We proceed by contradiction: suppose that $x_t^e \geq x_t^o$. Then:

$$R(x_{t-1}^o, x_t^o - x_{t-1}^o) \geq R(x_{t-1}^e, x_t^o - x_{t-1}^e) > R(x_{t-1}^e, x_t^e - x_{t-1}^e)$$

The first inequality holds because, for fixed next period's capacity $x + y$, $R(., .)$ is increasing in x. The second inequality follows from the monotonicity of $R(., .)$ in its second argument. By the equilibrium condition, $R(x_{t-1}^e, x_t^e - x_{t-1}^e) = c$ and together with the above inequality, we have $R(x_{t-1}^o, x_t^o - x_{t-1}^o) > c$ which contradicts (6) since the last two terms in the RHS of (6) are positive. Thus, $x_t^e < x_t^o$.

Note that the discrepancy between the equilibrium and the optimal paths we identify here is compatible with Schumpeter's [1950] argument that competitive markets may be not dynamically optimal, because small firms cannot appropriate the returns of their investments.

The case of negative externalities. The case of negative demand feedbacks ($m < 0$) corresponds to situations where the willingness of the consumers to pay for the good is decreasing in the volume of past sales. Reasons for this pattern of behavior include receiving negative information from previous buyers about the characteristics of the product, or what is often referred to as the "snob effect."

For $m < 0$, there can be no unambiguous comparison of the two paths. The efficient path may be either above or below the equilibrium path. There are two competing effects. On the one hand, the planner expands capacity more quickly because learning about the potential demand, θ, is internalized. On the other hand, for $m < 0$, the planner wants to expand less quickly, because today's expansion decreases future demand. These effects can be demonstrated by a comparison of (3) and (6). The second term of the RHS of (6) is positive, whereas the last term is negative. Which of the two effects dominates depends on the relative strength of the parameters.

Further, capacity expansion for the planner may *stop* at a price level at which entry by the competitive market would have continued. Recall that competitive entry would stop at time t if and only if $p_t < c(1 - \beta)$. Clearly, at such a price, expansion by the planner

will also stop. The reason is that, since with $m < 0$ the price never increases, further capacity expansion will always decrease profits (the loss would be $c - p_t/(1-\beta) > 0$ for each additional unit, and the profitability of the already invested units will decrease, since the price decreases). However, unlike the competitive case, the planning path may also stop at some level x_s with price p_s *higher* than $c(1-\beta)$. The reason is that the planner recognizes the negative effect that further expansion will have on the profitability of the previously installed capacity. Thus, it is possible that the optimal strategy is to stop, with future profit equal to $p_s x_s/(1-\beta)$. Even without considering the risk of exceeding the demand for one period, increasing sales will decrease the price for *all* units of subsequent sales. Thus, the per unit profit for the additional units has to be compared with the decreased profitability of the installed capacity. If $m < 0$ is low enough, x_s is large enough, and β is large enough (so that future profits have enough value), then it is optimal to never increase capacity from the level x_s.

A26. **Proof of Proposition 4: Entry with low initial demand**. As discussed in the text, clearly there in no entry in the competitive case. Now we move to optimal entry. (i) Assume that there is investment x_1 in the first period. If $x_1 < \theta$, the firm enters period two with positive updating about the demand, θ. It has also raised the price from p_1 to $p_2 = p_1 + mx_1 > p_1$. Thus, if the firm has decided to invest in the first period, it must be optimal for it to invest in the second period, and so on. It cannot be optimal for the firm to decide to have losses in the first period and then stop when demand turns out to be high.

(ii) Given (i) and the distribution F, the firm can calculate the expected profit from entering the market. By construction, the optimal path satisfies (6). The firm will invest if and only if the expected present value of profits is positive; otherwise investment never takes place. Furthermore, the value of the optimal expansion path is increasing and continuous (for continuous F) in m. For $m = 0$, this value is negative (since $p_t = p_1 < c(1-\beta)$, for all t). As $m \to \infty$, on the other hand, the value becomes very large. It follows that there is a unique $\bar{m}$, $0 < \bar{m} < \infty$, such that the value of the project is exactly zero. Investment takes place if and only if $m > \bar{m}$.

The planner is not cash constrained. Thus, the planner can accumulate losses for some time with the prospect of future profits. Otherwise, there is an additional constraint to consider and the set of markets that it is optimal to enter becomes smaller. On average, stricter finance constraints imply that a higher slope, m, is required for the project to be undertaken. In addition, taking into consideration finance constraints may change the expansion path even for a market that it is still optimal to enter. The reason is that, even when the optimal solution for the unconstrained problem is not feasible in the constrained problem, the optimal solution to the constrained problem may still yield positive profit (that

10

is, higher than the profit from not entering, which is zero). It may be interesting to further examine, in this framework, the role of finance constraints on investment in new markets.

A30. **Policy implications and infant export industries**. Further, consistency in the treatment of established and infant industries was central in this policy: promotion would end once either the industry matured or it failed to "succeed" within a period of time (in which case the initial decision to support the industry was reversed on the basis of "new information" that had been obtained along the way). The same discussion views dynamic externalities among firms as playing a central role in shaping industrial policies. In particular, it was recognized that intra-industry externalities are important and "due to the fact that one agent's investments to obtain information can very significantly reduce transaction costs for access by other nearly agents to the same and closely related information." (Pack and Westphal [1986], p. 110).

A31. **Proof of Proposition 5: The competitive path is S-shaped for m large enough**. To prove the Proposition we first prove the following.

Lemma. The diffusion path is concave if $y_2 \leq y_1$. If $y_2 > y_1$ the path is S-shaped (unless θ is low enough that entry stops while it is still increasing).

The proof of the Lemma is as follows. Suppose that $y_t \leq y_{t-1}$ (or, equivalently, $x_t - x_{t-1} \leq x_{t-1} - x_{t-2}$). Then,

$$\Delta p_t = p_{t+1} - p_t = m(x_t - x_{t-1}) \leq m(x_{t-1} - x_{t-2}) = p_t - p_{t-1} = \Delta p_{t-1}$$

Thus, if $y_2 \leq y_1$, as t increases the "hazard rate" effect becomes stronger (as capacity increases), and the "demand shifting" effect becomes weaker ($\Delta p_t \leq \Delta p_{t+1}$), and therefore y_t is decreasing and the path is concave.

If $y_1 < y_2$ the path is initially convex, by construction. Eventually concavity sets in, as capacity increases and the hazard rate becomes high enough. Thus, there is a $T > 2$ such that $\Delta y_t > 0$ for $2 \leq t \leq T$ and $\Delta y_t \leq 0$ for $t > T$, that is, the diffusion curve is S-shaped. Of course, θ may be low so that starting from $y_1 < y_2$, the path is convex until entry stops. Note that when $y_2 \geq y_1$, then $\Delta p_2 \geq \Delta p_1$ but y_3 may be either larger or smaller than y_2. This completes the proof of the Lemma.

Further, it is easy to see that if $\bar{\theta} \leq 2x_1$, then $y_2 < y_1$, and, by the above Lemma, the path is necessarily concave. Using (4), $\bar{\theta} \leq 2x_1$, is equivalent to

$$2c(\bar{\theta} - \underline{\theta}) - \bar{\theta}(p_1 + \beta c) \leq 0.$$

Thus we assume $2c(\bar{\theta} - \underline{\theta}) - \bar{\theta}(p_1 + \beta c) > 0$ in Proposition 5, otherwise the path has no chance to be S-shaped.

11

To prove Proposition 5, note that by the above Lemma, it is enough to show that $m >$ $m^* \Leftrightarrow y_2 > y_1$. But $y_2 > y_1 \Leftrightarrow x_2 - x_1 > x_1 \Leftrightarrow x_2 > 2x_1 \Leftrightarrow$

$$\bar{\theta} - \frac{c^2(\bar{\theta} - \underline{\theta})}{(p_2 + \beta c)(p_1 + \beta c)} > 2\bar{\theta} - \frac{2c(\bar{\theta} - \underline{\theta})}{(p_1 + \beta c)} \Leftrightarrow p_2 \geq \frac{c^2(\bar{\theta} - \underline{\theta})}{2c(\bar{\theta} - \underline{\theta}) - \bar{\theta}(p_1 + \beta c)} - \beta c.$$

Now, $p_2 = p_1 + mx_1$ with x_1 given by (4) and substituting accordingly we obtain $m > m^*$ Direct calculations, keeping in mind that $p_1 > c(1 - \beta)$, show that $m^* > 0$.

A32. The shape of the optimal expansion path.

First note that we know that $y_1^o > y_1^e$. When we compare the optimal and the equilibrium paths at $t = 2$ we find three differences. First, since $y_1^o > y_1^e$ the hazard of overshooting is higher for the optimal path; thus there is a tendency for $y_2^o < y_2^e$. Second, the investment incentives that were present at $t = 1$ are still present at $t = 2$ and this creates a tendency for $y_2^o > y_2^e$. Third, $y_1^o > y_1^e$ also implies higher willingness of the consumers to pay at $t = 2 : p(y_1^o) > p(y_1^e)$. This last factor also creates an incentive for $y_2^o > y_2^e$, and it is the reason why the optimal path can be, in principle, S-shaped even when the competitive path is not.

We now examine how the parameters affect the time path of investment. First, for $\beta = 0$ the competitive equilibrium and the optimal paths are *identical*. This comes from equation (6) and should also be intuitively clear because with $\beta = 0$ investment issues (in learning or market cultivation) become irrelevant. By continuity of the optimal solution we conclude that for β close to 0 and for parameter values that make the equilibrium path S-shaped (that is, $m > m^*$; see Proposition 5), the optimal path will be S-shaped as well. As β increases towards 1, the incentives for investment become higher (since the present value of the rewards is higher), and x_1^o increases relative to investments in subsequent periods. It follows that, higher values of β make it harder for optimal paths to be S-shaped. Now we turn to m. If $m = 0$ (fixed demand) then (7) becomes linear and the stable manifold is the saddle path itself (see the analysis of the dynamical system in topic A23 above). Thus if $m = 0$, the optimal solution is equation (A23.3) (see topic A23) which implies that the optimal entry path is monotonically decreasing (that is, the expansion path is concave). For values of m close to 0 the solution will also be concave as well (by continuity). As m increases it is possible, for some parameter values, that $y_1^o < y_2^o$ and the path is S-shaped. However, as m increases we find that for high enough values of m, x_1^o becomes very high so that the path is necessarily concave. Formally, this can be shown as follows. It is easy to verify (using (A23.5)) that $\bar{z} \to \bar{\theta}$ as $m \to +\infty$, which since $\bar{z} \leq x_1^o \leq \bar{\theta}$, implies that x_1^o is close to $\bar{\theta}$ for m high enough. Thus, the optimal path is concave for low and for high values of m and can be S-shaped for intermediate values.

12

Finally, the fact that the optimal path cannot be outside of area A (see the topic A23 above and Figure A4) can be used to provide some more precise information. For example, since $x_o = 0$ we know that $x_1^o \geq \bar{z}$ (where $\bar{z}$ is defined in (A23.5)). We also know that $x_2^o < \bar{\theta}$. Thus, since $y_2^o \geq y_1^o \Leftrightarrow x_2^o \geq 2x_1^o$, we know that if $\bar{\theta} < 2\bar{z}$ then $y_2^o < y_1^o$. Using the equation for $\bar{z}$, it is easy to show that $\bar{\theta} < 2\bar{z} \Leftrightarrow 2(p + \beta c) + m\bar{\theta}\beta > 4c$. Thus, we have that if $2(p + \beta c) + m\bar{\theta}\beta > 4c$ the optimal path satisfies $y_2^o < y_1^o$.

A33. **Entry and di¤usion paths**. We discuss here briefly some regularities that have been empirically documented in new markets and are related to our analysis. Before proceeding to the discussion, note that, in many new markets, technology changes heavily influence the dynamics. Our analysis abstracts away from technology improvement issues to focus on informational concerns and demand endogeneity. Technology factors have been studied in a number of papers (see e.g. Jovanovic and Lach [1989], Jovanovic and MacDonald [1994], and Klepper [1996].) The analysis here should be viewed as a complement, not a substitute for this work. In markets where both demand endogeneity and technology factors are important, the insights from both approaches have to be combined.

There are a number of regularities related to our analysis (see e.g. Gort and Klepper [1982], Dunne, Roberts and Samuelson [1988], Klepper and Graddy [1990], Klepper [1996], Geroski [1995], and references therein). First, entry and capacity expansion are gradual, not instantaneous. Second, total invested capacity in the industry grows over time until it reaches a maximum level (when a "shake-out" stage starts). In our framework, firms enter over time as new information about the demand becomes available. Third, following the initial entry into the market, entry rates are often increasing over a period of time and then start decreasing or, in other cases, achieve a maximum at the time of initial entry and decrease afterwards. In either case, eventually entry rates become small (see e.g. Figure 1 in Klepper [1996], p. 564). These two entry patterns imply cumulative expansion paths that are either S-shaped or concave.

A36. **Epidepic-type models of di¤usion**. The marketing literature has provided a number of studies on the diffusion of new products. This literature has been largely based on an epidemic-type approach, as formulated by Bass [1969]. Central in this approach are two parameters, the coefficient of "innovation" and of "imitation." A larger coefficient of imitation may imply that early purchases influence later purchases to a larger extent, and comparing such coefficients could be suggestive of differences across markets. Note, however, that a key difference from our work is that the Bass [1969] class of models starts with a specification of these relations at the level of sales. Instead, our approach takes as given only the demand and it is thus possible, for example, to derive different predictions for the paths in a competitive or a monopolistic market.

A39. **Empirical studies and extensions**. When taking the basic ideas to the data, the framework needs to be extended to allow for aspects of reality not present in the model. Two such important dimensions are mentioned here. First, there is a body of work demonstrating that firm heterogeneity, exit, and technology improvement are important in industry dynamics. Firms differ with respect to their technology and information, selection takes place leading to less efficient firms dropping out, and products become less costly to produce. Second, demand endogeneity is captured in the model in a simple parametric way but may take a number of different forms in reality. For example, the formal model assumes that past sales increase the consumers' willingness to pay but the potential demand, θ, remains unaffected. Assuming that we also have an increasing function $\theta(x_{t-1})$ would leave the main results qualitatively unaffected as long as, in addition to the growth component, the market size also had a factor that firms learn about.

We also discuss here two other possible extensions. First, it is worth examining more general demand structures that imply different monopoly and socially optimal paths. Clearly, the model presented here overemphasizes the benefits from concentration since, by construction, the monopoly and the planning problems coincide. In other words, the issue here is why, due to dynamic considerations and the presence of externalities, the competitive market will expand at a rate that is too low. Before proceeding to policy recommendations one has to also consider the standard monopoly distortions that would tend to make a monopolist produce at suboptimal levels. In a more general model, an interesting trade-off would appear. More concentrated industries (monopolies, in the extreme case) would tend to internalize the externalities and invest more, but would also tend to produce less than the optimum in later stages. One can conjecture that higher concentration may be desirable for export-oriented industries (where consumers' welfare considerations do not arise) but not for industries that are primarily oriented to the domestic market. A second extension is the examination of investment paths when there is more than one competing product (such as different technological standards). Then tomorrow's demand for a given product would depend positively on past sales of the same product but negatively on the past sales of the competing products. In addition, when deciding to invest in this market, firms have demand uncertainty that is partly due to the behavior of their competitors. Several issues arise related to the optimal number and network-size of competing products, the pattern of entry and prices, and path-dependence.

- *A list of references for this Appendix follows.*

References

[1] Frank M. Bass. A new product growth model for consumer durables. *Management Science*, 15:215–227, 1969.

[2] Barry L. Bayus. High-definition television: Assessing demand forecasts for a next generation consumer durable. *Management Science*, 39:1319–1333, 1993.

[3] Timothy Dunne, Mark J. Roberts, and Larry Samuelson. Patterns of firm entry and exit in U.S. manufacturing industries. *Rand Journal of Economics*, 19:495–515, 1988.

[4] Joseph Farrell and Carl Shapiro. Standard setting in high-definition television. *Brookings papers: Microeconomics*, pages 1–93, 1992.

[5] Paul A. Geroski. What do we know about entry? *International Journal of Industrial Organization*, 13:421–440, 1995.

[6] Michael Gort and Steven Klepper. Time paths in the diffusion of product innovations. *Economic Journal*, 92:630–653, 1982.

[7] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springler-Verlag, 3rd edition, 1983.

[8] Boyan Jovanovic and Saul Lach. Entry, exit and diffusion with learning by doing. *American Economic Review*, 79:690–699, 1989.

[9] Boyan Jovanovic and Glenn MacDonald. The life cycle of a competitive industry. *Journal of Political Economy*, 102:322–347, 1994.

[10] Steven Klepper. Entry, exit, growth and innovation over the product life cycle. *American Economic Review*, 86:562–583, 1996.

[11] Steven Klepper and Elizabeth Graddy. The evolution of new industries and the determinants of market structure. *Rand Journal of Economics*, 21:27–44, 1990.

[12] Howard Pack and Larry E. Westphal. Industrial strategy and technological change: Theory versus reality. *Journal of Development Economics*, 22:87–128, 1986.

[13] Rafael Rob. Learning and capacity expansion under demand uncertainty. *Review of Economic Studies*, 58:655–675, 1991.

[14] Joseph Schumpeter. *Capitalism, Socialism and Democracy*. Harper and Row, New York, 3rd edition, 1950.

[15] Nikolaos Vettas. Demand and supply in new markets: Diffusion with bilateral learning. *Rand Journal of Economics*, 29:215–233, 1998.
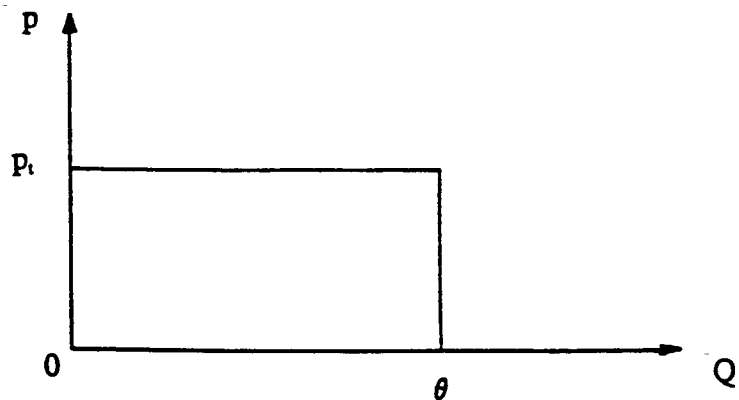
FIGURES
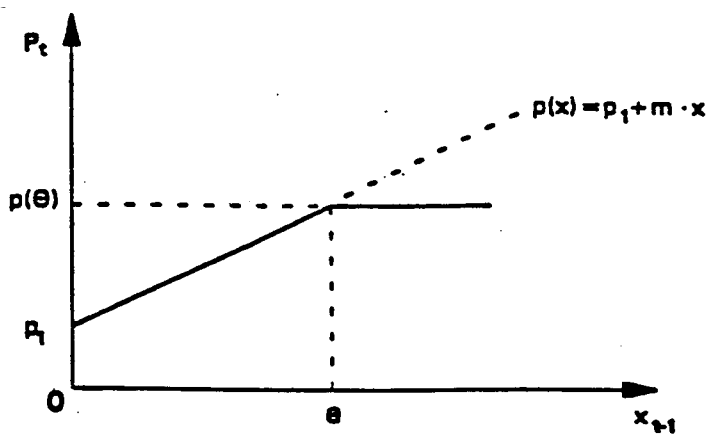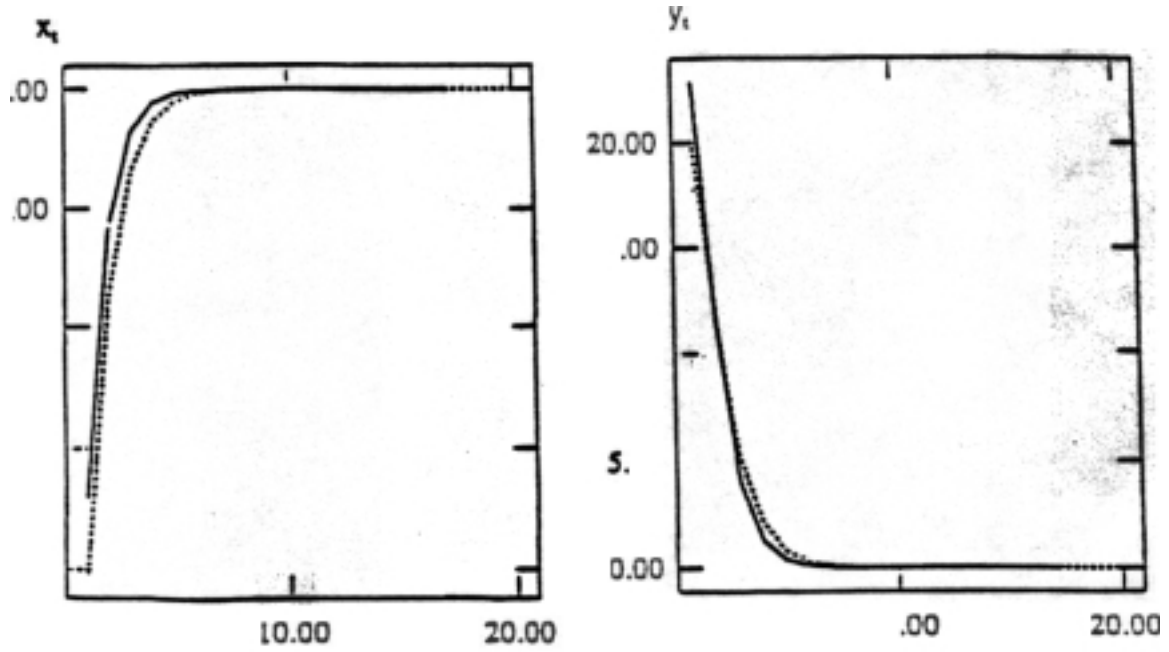

Figure A1a. The demand function in period t.

Figure A1b. p_t as a function of capacity x_{t-1} in the previous period.

1442
0.2442

Figure A2 Examples of the optimal path for
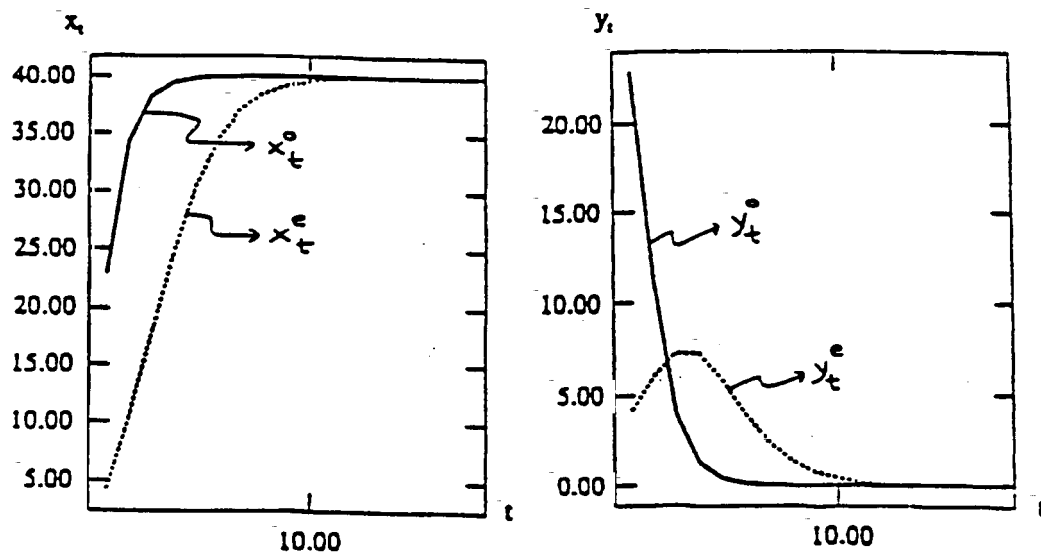β = 0.95, F uniform $[0, 40]$

Figure A3. Comparison of the equilibrium (y_t^e and x_t^e) and the optimal (y_t^o and x_t^o) paths for $p_1 = 2, c = 12, \beta = 0.95, F$ uniform on $[0, 40], m = 0.2442 > m^* = 0.0442$.
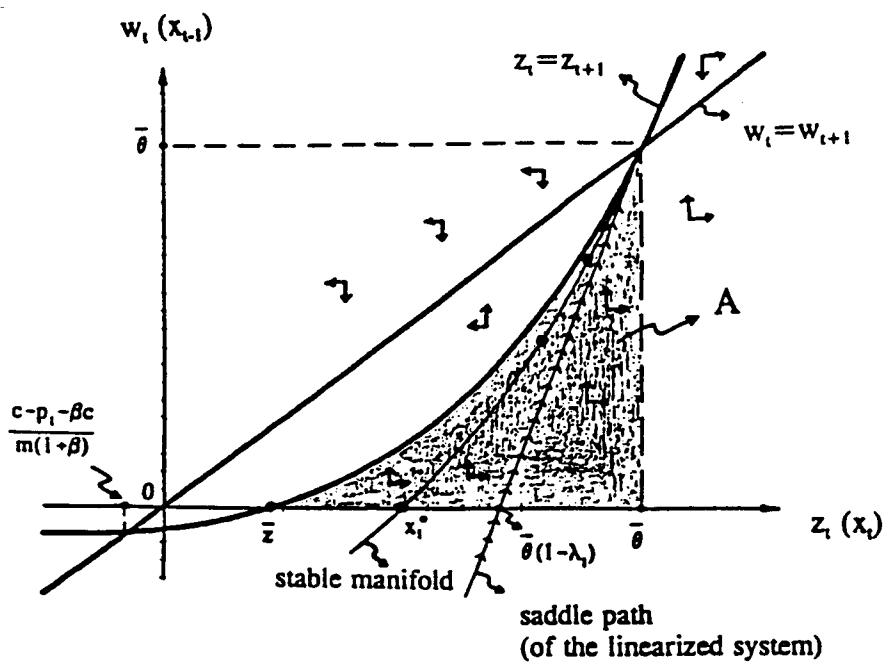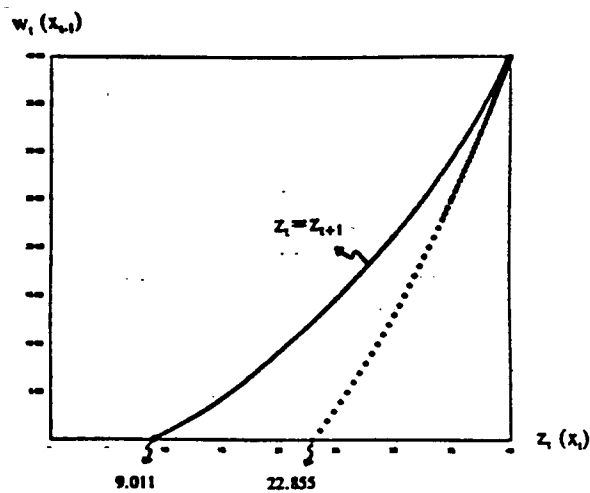
Figure A4. The dynamical system.

Figure A5. Paths coverging to $\bar{\theta} = 40$ for different initial values $x_0 = 0, 1, 2, ..., 40$ and for $p_1 = 2, c = 12, \beta = 0.95, F$ uniform on $[0, 40]$, $m = 0.2442$. All points are on the stable manifold.

t	x_t^o	x_t^e
1	22.85662	4.179104
2	34.18828	10.19215
3	38.20163	17.48848
4	39.45981	24.71332
5	39.83921	30.56189
6	39.95226	34.57179
7	39.98584	37.01801
8	39.99580	38.40546
9	39.99875	39.16004
10	39.99963	39.56108

t	y_t^o	y_t^e
1	22.8566	4.17910
2	11.3316	6.01305
3	4.01335	7.29633
4	1.25818	7.22483
5	0.37939	5.84856
6	0.11305	4.00990
7	0.03357	2.44621
8	0.00995	1.38744
9	0.00295	0.75458
10	0.00087	0.40104

Table A1. Comparison of the equilibrium (y_t^e and x_t^e) and the optimal (y_t^o and x_t^o) paths for $p_1 = 2, c = 12, \beta = 0.95$, F uniform on $[0,40]$, and $m = 0.2442 > m^* = 0.0442$.

t	y_t	x_t	$\pi_t = p_t x_t - c y_t$	$\sum_{\tau-1}^{t} \pi_t$
1	19.9961	19.9961	-1159.77939	-1159.77939
2	11.9401	31.9363	-13.9311261	-1173.71052
3	5.11956	37.0559	950.368949	-223.341573
4	1.90955	38.9654	1407.25879	1183.91722
5	0.67596	39.6414	1583.37237	2767.28960
6	0.23488	39.8763	1646.41461	4413.70421
7	0.08108	39.9574	1668.40401	6082.10823
8	0.02793	39.9853	1676.00546	7758.11370
9	0.00961	39.9949	1678.62507	9436.73877
10	0.00331	39.9982	1679.52687	11116.2656

Table A2. Example of an optimal expansion path with $p_1 < c(1 - \beta)$.
$p_1 = 2, c = 60, \beta = 0.95, F$ uniform on $[0, 40], m = 1$.