4

Expandability, Reversibility, and Optimal Capacity Choice

AVINASH K. DIXIT
ROBERT S. PINDYCK

We develop continuous-time models of capacity choice when demand fluctuates stochastically, and the firm has limited opportunities to expand or contract. Specifically, we consider costs of investing or disinvesting that vary with time, or with the amount of capacity already installed. The firm's limited opportunities to expand or contract create call and put options on incremental units of capital; we show how the values of these options affect the firm's investment decisions.

4.1 INTRODUCTION

Our recent book and survey articles on the real options approach to investment identify three characteristics of most investment decisions: (1) uncertainty over future profit streams, (2) irreversibility, i.e., the existence of some sunk costs that cannot be recouped if the firm changes its mind later, and (3) the choice of timing, i.e., the opportunity to delay the investment decision.[1] We argued that because of the *interaction* of these three forces, optimal investment decisions have to satisfy more stringent hurdles for their expected rates of return than the naive net present value (NPV) criterion would indicate. The uncertainty implies that there may be future eventualities where the firm would regret having invested. The irreversibility implies that if the firm invests now, it cannot costlessly disinvest should such an eventuality materialize. And the opportunity to wait allows it to learn more about the uncertain future and reduce the likelihood of such regret.

By analogy with financial options, the opportunity to invest is a call option—a right but not an obligation to make the investment. To invest is to exercise the option. Because of the uncertainty, the option has a time premium or holding value: it should

---

[1] See Dixit (1992), Pindyck (1991), and Dixit and Pindyck (1996).

not be exercised as soon as it is "in the money," even though doing so has a positive NPV. The optimal exercise point comes only when the option is sufficiently "deep in the money," i.e., the NPV of exercise is large enough to offset the value of waiting for more information. This conclusion is probably the most widely known "result" of the real options literature.

Of the triad of conditions mentioned above, most of the literature has focused on irreversibility. But most formal models assume simultaneously total irreversibility and a completely costless ability to wait, so they cannot separately identify the contributions of these two conditions. Exceptions to this include the seminal article by Brennan and Schwartz (1985), which examined an investment in a mining project and allowed for both an option to invest and an abandonment option, the models developed by Trigeorgis (1993, 1996) that allow for a variety of different options interacting within a single project, including options to expand and contract, and the work of Kulatilaka (1995) on substitutability and complementarity in real options. Another exception is our recent article co-authored with Abel and Eberly (1996), which developed a two-period model that allowed for arbitrary degrees of irreversibility and future expandability.

Abel, Dixit, Eberly, and Pindyck (henceforth referred to as ADEP) showed that a firm that makes an investment that is partially or totally reversible acquires a put option, namely the ability to pull out should future conditions be sufficiently adverse. This option has value if future uncertainty involves a sufficiently large downside with a positive probability that the firm will want to exercise the option. Recognition of this put option will make the firm *more* willing to invest than it would be under a naive NPV calculation that assumes that the project continues for its physical lifetime and omits the possibility of future disinvestment.[2] Likewise, a firm that can expand capacity by making an investment now or in the future (at a specified cost) is exercising a call option, acting now when it might have waited. This option has value if future uncertainty has a sufficiently large downside that waiting would have been preferable. Recognition of this call option will make the firm *less* willing to invest than it would be under a naive NPV calculation that assumes that the project must be started now or never, ignoring the possibility of a future optimal startup decision.

For many real-world investments, both of these options exist to some degree. Firms typically have at least some ability to expand their capacity at a time of their choosing, and sometimes can partially reverse their decisions by selling off capital to recover part of their investment. The net effect of these two options is in general ambiguous, depending on the degrees of reversibility and expandability, and the extent and nature of the uncertainty.

If the investment is totally irreversible, i.e., there is only a call option and no put option, the investment must necessarily satisfy a stiffer hurdle than a positive NPV (naively calculated). But as ADEP pointed out, it is not the irreversibility that gives rise to the call option; it is the *expandability* that does so. What irreversibility does is to eliminate the put option that acts in the opposite direction.

---

[2] The option to abandon a project midstream is an example of this. Myers and Majd (1984) showed how this option can be valued as an American put option and its implications for the investment decision.

One might argue that in practice irreversibility is often more important than limited expandability, in part because of "lemons effects," but mostly because many unpredictable shocks are industry specific.[3] However, expandability can also be limited, e.g., because of limited land, natural resource reserves, or because of the need for a permit or license, only a limited number of which are being issued. Thus it is important to recognize and clarify the effects of these different underlying economic conditions.

In this chapter, we move beyond the two-period analysis in ADEP to examine a set of continuous-time models that allows for incremental capacity expansion and/or contraction over time, and thereby provides further insight into the effects of irreversibility, expandability, and the ability to wait. In this continuous-time setting, limited reversibility and expandability lead to clearly identifiable (and measurable) put and call options, which have opposite effects on the firm's incentive to invest.

Most of our analysis deals with exogenous and time-dependent limitations on the firm's ability to expand or contract. Specifically, we consider models in which the cost of investing increases over time (limited expandability) and the price that the firm can get by selling previously installed capital declines over time (limited reversibility). Our general framework is described in Section 4.2. In Section 4.3, models with time-varying costs are presented in detail and their implications for investment and capacity choice are examined. Section 4.4 examines the static and dynamic effects of sunk costs. In Section 4.5 we briefly discuss capacity choice decisions when the cost of investing or disinvesting varies with the amount of capacity already installed. In the concluding section, we suggest some extensions of our model for future work.

4.2 CONTINUOUS-TIME MODELS OF CAPACITY CHOICE

The two-period model developed in ADEP showed the effects of the call and put options associated with investment in the simplest possible way. For a more realistic analysis, however, we need a longer horizon, with ongoing uncertainty and repeated opportunities for the firm to expand or contract in response to changing circumstances. In such a setting, partial reversibility and expandability will arise when the costs of capacity contraction or expansion vary in response to changes in one or more exogenous or endogenous variables. In this chapter we consider two such variables.

First, we examine what happens when the cost of investing or disinvesting varies exogenously with time. This would be the case, for example, if the cost of capacity expansion rises over time as the resources needed for expansion (e.g., land or mineral reserves) are used up by other firms or dwindle for physical reasons (such as land erosion or the depletion of a potentially discoverable resource base). Likewise, the resale price of used capital is likely to fall over time, partly as a result of the increasing obsolescence of capital.

---

[3] For example, a steel manufacturer will want to sell a steel plant when the steel market is depressed, but that is precisely the time when no one else will want to pay a price for it anywhere near its replacement cost. Therefore investment in a steel plant is largely irreversible.

Second, we examine investment decisions when the cost of investing varies endogenously with the amount of capacity already installed by the firm. This kind of limited expandability would arise when the firm itself (which presumably has some monopoly power) uses up limited resources as it expands.

In both cases we assume that the firm faces an isoelastic demand curve of the form

$$P = \theta(t)Q^{-1/\eta} \tag{4.1}$$

where P is the price, Q is the quantity demanded, η is the elasticity of demand, and the demand shift variable θ varies stochastically according to the geometric Brownian motion:

$$d\theta = \alpha\theta \, dt + \sigma\theta \, dz \tag{4.2}$$

Although this is not critical, we assume for convenience that the uncertainty over future values of θ is spanned by the capital markets. Hence there is some risk-adjusted rate of return for θ, which we denote by μ, that allows risk-free discounting. We let $\delta = \mu - \alpha$ denote the rate-of-return shortfall.

To simplify matters, we assume that the firm has zero operating costs and hence will always produce at capacity, denoted by K. This eliminates any "operating options" that can affect the value of a unit of installed capital, allowing us to focus exclusively on options associated purely with the investment or disinvestment decision.[4]

As in Pindyck (1988), we examine the firm's incremental investment decisions. Let $\Delta V(K; \theta, t)$ denote the value of the last incremental unit of installed capital, and let $\Delta F(K; \theta, t)$ denote the value of the firm's option to install one incremental unit. In the standard neoclassical model of investment, ΔV would simply be the present value of the expected flow of marginal revenue from the unit in perpetuity, i.e.,

$$\Delta V_0(K; \theta, t) = \omega(K)\theta \tag{4.3}$$

where

$$\omega(K) \equiv \left(\frac{\eta - 1}{\eta\delta}\right) K^{-1/\eta} \tag{4.4}$$

Likewise, ΔF is the greater of zero or the NPV of immediate investment in this incremental unit. If the cost of an incremental unit of capital were fixed at k_0, then ΔF in the neoclassical model would be given by

$$\Delta F_0(K; \theta, t) = \max[0, \ \omega(K)\theta - k_0] \tag{4.5}$$

---

[4] The most important operating option is the ability of the firm to reduce output or to shut down and thereby avoid variable operating costs. As demonstrated by McDonald and Siegel (1985), this operating option raises the value of a unit of capital. For a discussion of this and other operating options, see Chapter 6 of Dixit and Pindyck (1996).

The neoclassical model, however, ignores the value of the firm's options to buy or sell capacity in the future. These option values depend on the firm's ability to make such purchases or sales, and the prices it will pay or receive for capital. In the next section, we allow those prices to vary with time.

4.3 TIME-DEPENDENT COSTS

Suppose that additional capacity can be added at a cost $k(t) = k_0 e^{\rho t}$ per unit, with $\rho \geq 0$. In this setting, if $\rho > 0$ so that the cost of adding capacity is rising over time, there is *limited expandability*; with $\rho = 0$ there is complete expandability; and for $\rho \to \infty$ there is no expandability. In this model, limits to expandability are exogenous to the firm's actions; $k(t)$ might rise, for example, because of continual entry or expansion by other firms that pushes up capital costs. Although we consider only values of $\rho \geq 0$, in practice ρ may in some cases be negative. This could occur, for example, if continual technological improvements or learning by doing cause per unit capital costs to fall over time.

Similarly, we assume that installed capital can be sold, but only for a price $S(t) = k_1 e^{-st}$ per unit, with $s \geq 0$. Hence there is *partial reversibility* that is completely time dependent, reflecting, for example, the increasing obsolescence of capital (as opposed to its physical depreciation). If $k_1 = k_0$, then at $s = 0$ investment is completely reversible. If $k_1 < k_0$, there is some irreversibility (even if $s = 0$); this can arise because of "lemons" effects. In either case, if $s \to \infty$, investment is completely irreversible. We call the effect of an initial gap between the purchase and sale prices of capital the "static" aspect of irreversibility, and the widening of the gap over time because of $\rho > 0$ and $s > 0$ the "dynamic" effect. At first we focus on the dynamic effect by assuming $k_1 = k_0$. In Section 4.4 we bring in the static effect, and compare the two.

At this point it is useful to explain our modeling choices. In making the cost of installing capital purely a function of time, we have in mind that cost increases are largely the result of the activities of other firms. For example, in an extractive industry such as oil or copper, other firms will deplete the potentially discoverable resource base over time; then expansion by a given firm becomes more expensive over time as new deposits are harder to find and costlier to develop. In the residential and commercial construction industries, other firms will buy and develop choice parcels of land over time, making expansion by a given firm more expensive. Ideally, this process should be modeled in an equilibrium setting, so that each firm in the industry (including possible new entrants) makes its decisions consistent with rational expectations of the optimal behavior of all other firms. Although equilibrium models of entry and exit with sunk costs are available in the literature [e.g., see Chapters 8 and 9 of Dixit and Pindyck (1996) for an overview], here we focus on the optimal decisions of the manager of one firm. Most managers base their decisions on expectations of changes in market parameters, including capital purchase and resale prices. Managers may or may not think in terms of an overall industry equilibrium when they form these expectations, but they often tend to treat these price movements as exogenous functions of time, much as we treat them here.

We offer a similar justification for our assumption that capital can be sold for a price that declines with time, irrespective of when the capital was purchased. Again, we have in mind a pattern of obsolescence that is largely caused by other firms that are continually developing superior processes and/or products.[5]

Finally, one might question our assumption that the purchase and sale prices of capital evolve *exponentially* with time, which was chosen for analytical convenience. Although other forms of time dependence might be introduced that may be more realistic for particular industries, this would complicate the algebra that follows.

With these caveats in mind, the basic idea we develop is that limited expandability and reversibility create options that must be taken into account when determining the firm's optimal investment rules. In contrast to the neoclassical model, ΔV actually has two components: the value of the expected profit flow from the use of the incremental unit of capital and the value of the (put) option to sell the unit in exchange for the salvage amount $k_0 e^{-st}$. Likewise, ΔF accounts for the full option value of the investment, i.e., the fact that the option has a time value and need not be exercised immediately.

Using standard methods, it is easy to show that ΔV must satisfy the following differential equation:

$$\tfrac{1}{2}\sigma^2\theta^2 \Delta V_{\theta\theta} + (r - \delta)\theta\Delta V_\theta + \Delta V_t - r\Delta V + \delta\omega(K)\theta = 0 \qquad (4.6)$$

where $\omega(K)$ is given by Eq. (4.4). The solution must also satisfy the following boundary conditions:

$$\lim_{\theta\to\infty} (\Delta V/\theta) = \omega(K) \qquad (4.7)$$

$$\Delta V(K; \theta^{**}, t) = \Delta F(K; \theta^{**}, t) + k_0 e^{-st} \qquad (4.8)$$

$$\Delta V_\theta(K; \theta^{**}, t) = \Delta F_\theta(K; \theta^{**}, t) \qquad (4.9)$$

Here $\theta^{**} = \theta^{**}(K, t)$ is the critical value of θ below which it is optimal to exercise the put option and sell the unit of capital. Boundary condition (4.7) simply says that if θ is very large, the firm will never want to sell off the unit of capital, so that its value is just the present value of the expected profit flow that it generates. Conditions (4.8) and (4.9) are the standard value matching and smooth pasting conditions that apply at the critical exercise point θ^{**}.

Likewise, ΔF must satisfy

$$\tfrac{1}{2}\sigma^2\theta^2 \Delta F_{\theta\theta} + (r - \delta)\theta\Delta F_\theta + \Delta F_t - r\Delta F = 0 \qquad (4.10)$$

subject to the boundary conditions:

$$\Delta F(K; 0, t) = 0 \qquad (4.11)$$

---

[5] Again, a fuller theory of such a pattern of technological "leapfrogging" might best be described in an equilibrium framework, but that goes beyond what we aim to do here. [Hence we are not considering physical depreciation, as in Chapter 6 of Dixit and Pindyck (1996), where the sale price begins declining only after the capital has been purchased.]

$$\Delta F(K; \theta^*, t) = \Delta V(K; \theta^*, t) - k_0 e^{\rho t} \qquad (4.12)$$

$$\Delta F_\theta(K; \theta^*, t) = \Delta V_\theta(K; \theta^*, t) \qquad (4.13)$$

$$\lim_{t \to \infty} \Delta F(K; \theta, t) = 0 \qquad (4.14)$$

Of these, the first three conditions are standard; the last one says that (with $\rho > 0$) the value of the call option to install an incremental unit of capital approaches zero as time passes, because the cost of exercising the option is rising exponentially.

To clarify the nature of the optimal investment decision, it is best to proceed in steps. As we noted in the introduction in Section 4.1, most of the literature assumes that investment is completely irreversible and completely expandable. We will begin by considering the case in which investment is completely irreversible but only partially expandable, so that there is only a single boundary, $\theta^*(K, t)$, which triggers investment. This special case helps to elucidate the nature of the call option and its dependence on the extent of expandability. In Section 4.3.2 we examine the case in which investment is partially reversible but completely nonexpandable, so that investment entails only a put option (the value of which depends on the extent of reversibility), but no call option. In this case there is again only a single boundary, $\theta^{**}(K, t)$, which triggers disinvestment. We return to the general case set forth above in Section 4.3.3.

4.3.1 Complete Irreversibility, Partial Expandability

In this special case $s = \infty$, so the firm cannot disinvest. Then ΔV is simply the present value of the flow of marginal revenue from an incremental unit of capital:

$$\Delta V(K; \theta, t) = \left(\frac{\eta - 1}{\eta \delta} \right) K^{-1/\eta} \theta = \omega(K) \theta \qquad (4.15)$$

We can find the solution to Eq. (4.10) for the value of the option to install an additional unit of capital by guessing a functional form and choosing its parameters to satisfy all of the boundary conditions.

We guess (and then verify) that the solution to Eq. (4.10) for ΔF has the form

$$\Delta F = a(K) \theta^{\beta_1} e^{-gt} \qquad (4.16)$$

The parameters β_1, g, and $a(K)$, along with the critical value θ^*, are found from the boundary conditions (4.11)–(4.14). By substituting Eq. (4.16) for ΔF into Eq. (4.10), we know that β_1 must be a solution to the fundamental quadratic equation

$$\tfrac{1}{2}\sigma^2 \beta_1(\beta_1 - 1) + (r - \delta)\beta_1 - r - g = 0 \qquad (4.17)$$

From condition (4.11), β_1 must be the positive solution to this equation, i.e.,

$$\beta_1 = \tfrac{1}{2} - \frac{(r - \delta)}{\sigma^2} + \sqrt{[(r - \delta)/\sigma^2 - \tfrac{1}{2}]^2 + 2(r + g)/\sigma^2} > 1 \qquad (4.18)$$

From conditions (4.12) and (4.13), the critical value θ^* is given by

$$\theta^*(K,t) = \left(\frac{\beta_1}{\beta_1 - 1}\right)\left(\frac{\eta\delta}{\eta - 1}\right) K^{1/\eta} k_0 e^{\rho t} \qquad (4.19)$$

Substituting this into boundary condition (4.12) gives the following expression for $a(K)$:

$$a(K) = (\beta_1 - 1)^{\beta_1 - 1}\left(\frac{\eta - 1}{\eta\delta\beta_1}\right)^{\beta_1} K^{-\beta_1/\eta} k_0^{1-\beta_1} e^{[g - \rho(\beta_1 - 1)]t} \qquad (4.20)$$

Since $a(K)$ cannot depend on t, $g = \rho[\beta_1(g) - 1]$. Substituting Eq. (4.18) for $\beta_1(g)$ gives

$$g = \rho\left[-\frac{1}{2} - \frac{(r - \delta - \rho)}{\sigma^2} + \right.$$
$$\left. \sqrt{[(r - \delta - \rho)/\sigma^2 + \frac{1}{2}]^2 + 2\delta/\sigma^2} \right] > 0 \qquad (4.21)$$

This solution for g and the relationship between g and β_1 can be seen more intuitively by rewriting (4.17) as

$$g = \frac{1}{2}\sigma^2\beta(\beta - 1) + (r - \delta)\beta - r$$

and plotting this along with the line $g = \rho[\beta - 1]$, as shown in Figure 4.1. The solution for g and β_1 is found at that intersection of these two curves at which $\beta > 1$.

Here, $\beta_1/(\beta_1 - 1) > 1$ is the standard "wedge" that arises in irreversible investment problems. But as $\rho \to \infty$, $\beta_1 \to \infty$ and $g \to \infty$. This can be seen algebraically from Eqs. (4.17) and (4.21) (or graphically from Figure 4.1) by observing that as ρ increases, the line $g = \rho(\beta - 1)$ twists counterclockwise around the point $(1,0)$. Then $\beta_1/(\beta_1 - 1) \to 1$, so that for $t = 0$, $\theta^* \to k_0/\omega(K)$, i.e., the value it
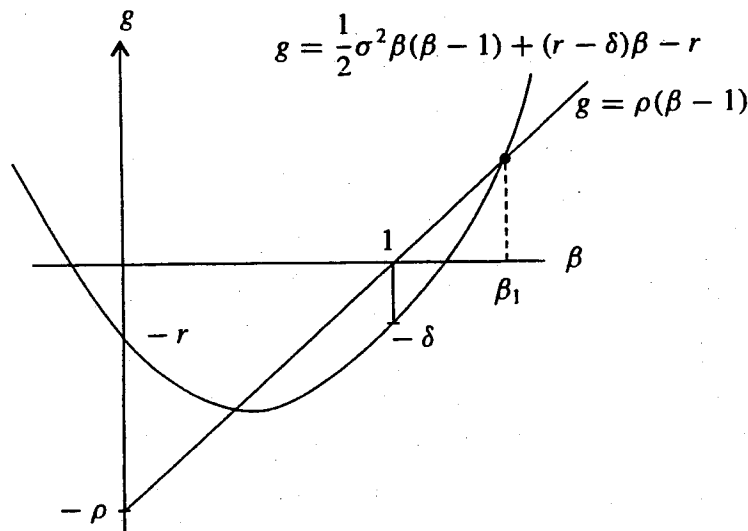

Figure 4.1 Complete irreversibility, partial expandability—graphs of Eq. (4.17) and the line $g = \rho(\beta - 1)$, providing solution for g and β_1.

would have in the absence of uncertainty. One can also see from Figure 4.1 that if $\rho \to \infty$, $\beta_1 \to \infty$ and $g \to \infty$, so that $\Delta F(K; \theta, t) = 0$ for $t > 0$, and

$$\Delta F = \max [0, \omega(K)\theta - k_0]$$

for $t = 0$. In this case ΔF is either zero or the net present value of the incremental investment—there is no option to invest after $t = 0$.

Figure 4.2 shows the optimal threshold $\theta^*(K, t)$, plotted as a function of capacity (K) for three values of t. (The parameter values are $r = 0.05$, $\delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, $\rho = 0.20$, and $k_0 = 3.0$.) Observe that the threshold boundary moves up over time as the cost of investing increases. Figure 4.3 shows $\theta^*(K, t)$ as a function of K at $t = 3$, for three different values of the volatility σ. As is typical in investment problems of this kind, the value of the call option increases as σ increases, and so does the threshold θ^* that triggers investment.
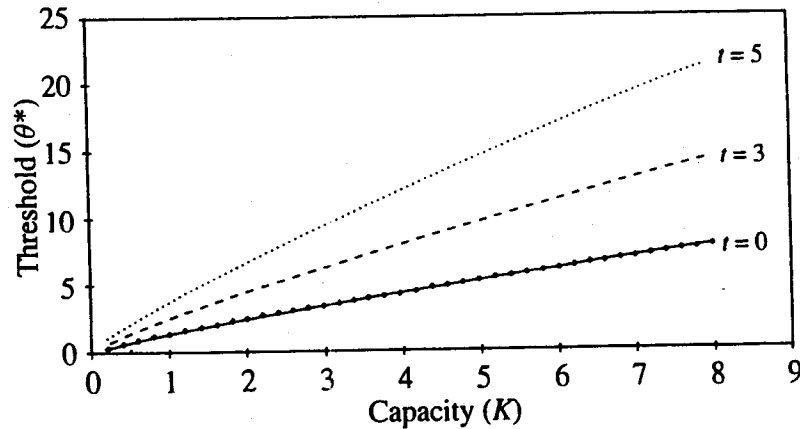

Figure 4.2   Complete irreversibility, partial expandability—demand threshold $\theta^*(K, t)$ as a function of capacity K. Parameter values: $r = 0.05$, $\delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, $\rho = 0.20$, and $k_0 = 3.0$.
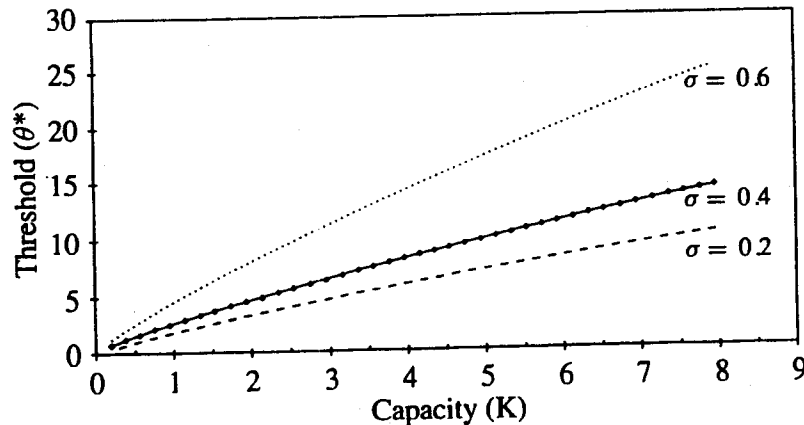

Figure 4.3   Complete irreversibility, partial expandability—dependence of investment threshold $\theta^*(K, t)$ on volatility σ.

Finally, if in addition to $s = \infty$ we let $\rho = 0$, we have the case that has received the most attention in the literature, namely complete irreversibility and complete expandability. In this special case, $g = 0$, β_1 is the solution to the standard quadratic equation, and θ^* becomes independent of time (see Dixit and Pindyck, 1996).

4.3.2 Partial Reversibility, No Expandability

This is the case for which $\rho = \infty$ and $s > 0$, so the firm can disinvest but cannot expand. Now the solution to Eq. (4.6) for ΔV is of the form

$$\Delta V(K; \theta, t) = b(K)\theta^{\beta_2}e^{-ht} + \left(\frac{\eta - 1}{\eta\delta}\right)K^{-1/\eta}\theta \qquad (4.22)$$

where the first term on the right-hand side is the value of the put option to sell the unit of capital. This solution can be verified by direct substitution in (4.6), and expressions for β_2, h, $b(K)$, and the critical value θ^{**} can be found using boundary conditions (4.7)–(4.9).

Investment can either occur immediately (at $t = 0$) or never, so ΔF is given by the standard NPV rule:

$$\Delta F(K; \theta, t) = \max[0, \ \Delta V(K; \theta, t) - k_0] \qquad (4.23)$$

for $t = 0$, and $\Delta F = 0$ for $t > 0$. In this case the boundary conditions that apply to ΔV are not linked to those for ΔF, so we can determine ΔV independently from ΔF. Since the firm has no call option to invest in the future, it will set its initial capacity K at the point where $\Delta V(K; \theta, 0) = k_0$. Hence the only issue is to determine ΔV.

Substituting Eq. (4.22) into (4.6) and using boundary condition (4.7), we find that β_2 is the negative solution to the quadratic Eq. (4.17), with g replaced by h, i.e.,

$$\beta_2 = \tfrac{1}{2} - \frac{(r - \delta)}{\sigma^2} - \sqrt{[(r - \delta)/\sigma^2 - \tfrac{1}{2}]^2 + 2(r + h)/\sigma^2} \ < 0 \qquad (4.24)$$

To obtain solutions for h, $b(K)$, and θ^{**}, we proceed as in the previous case, using boundary conditions (4.8) and (4.9) and the fact that $b(K)$ must be independent of t:

$$\theta^{**}(K, t) = \left(\frac{\beta_2}{\beta_2 - 1}\right)\left(\frac{\eta\delta}{\eta - 1}\right)K^{1/\eta}k_0e^{-st} \qquad (4.25)$$

$$b(K) = -\frac{1}{\beta_2}\left(\frac{\beta_2 - 1}{\beta_2}\right)^{\beta_2 - 1}\left(\frac{\eta - 1}{\eta\delta}\right)^{\beta_2}K^{-\beta_2/\eta}k_0^{1-\beta_2} \qquad (4.26)$$

and

$$h = s\left[\tfrac{1}{2} + \frac{(r - \delta + s)}{\sigma^2} + \sqrt{[(r - \delta + s)\sigma^2 + \tfrac{1}{2}]^2 + 2\delta/\sigma^2}\right] > s \qquad (4.27)$$

Note that $0 < \beta_2/(\beta_2 - 1) < 1$. As σ increases, β_2 increases toward 0, so that this multiple becomes smaller in magnitude. Thus the more uncertainty there is, the lower is the critical value of θ that will trigger disinvestment. This is a standard result [see

the model of entry and exit in Dixit (1989) or Chapter 7 of Dixit and Pindyck (1996)], but now this multiple depends on s, the rate at which the resale value of capital is falling. The larger is s, the closer this multiple is to one, and the smaller is $b(K)$ and hence the value of the put option. As $s \to \infty$, $\beta_2 \to -\infty$, and $\beta_2/(\beta_2 - 1) \to 1$; then there is no put option, so that for $t = 0$, $\theta^{**}(K) \to k_0/\omega(K)$, which is the value it would have in the absence of uncertainty.

Figure 4.4 shows the solution for the critical disinvestment threshold $\theta^{**}(K, t)$, again plotted as a function of capacity (K) for three values of t. (As before, the parameter values are $r = 0.05$, $\delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, and $k_0 = 3.0$, and now $s = 0.20$.) Observe that the boundary moves down over time as the price that the firm can receive for installed capital decreases. Figure 4.5 shows $\theta^{**}(K, t)$ as a function of K at $t = 3$, for three different values of σ. Since the value of the firm's put option increases as σ increases, the threshold that triggers disinvestment moves down.
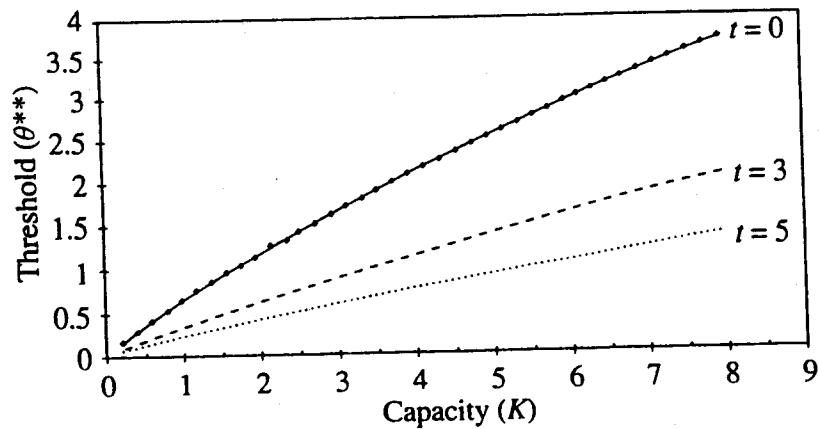

Figure 4.4 Partial reversibility, no expandability—disinvestment threshold $\theta^{**}(K, t)$ as a function of capacity K. Parameter values: $r = 0.05$, $\delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, $k_0 = 3.0$, and $s = 0.20$.
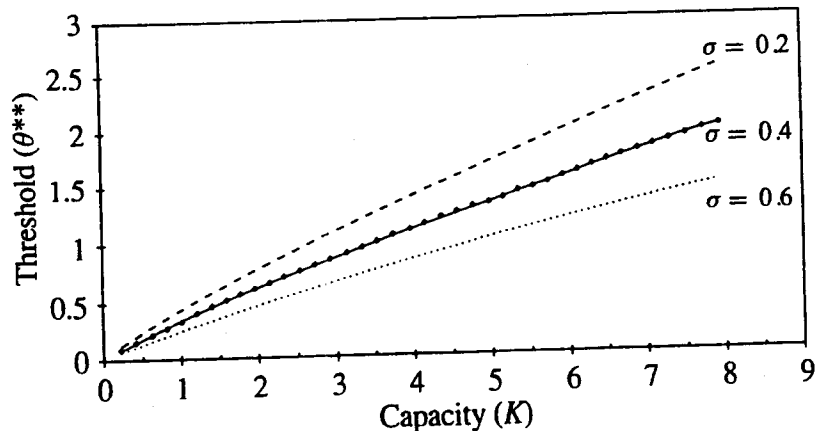

Figure 4.5 Partial reversibility, no expandability—dependence of disinvestment threshold $\theta^{**}(K, t)$ on volatility σ.

4.3.3 The General Case with Expandability and Reversibility

In the general case, both ρ and s are positive and finite. Now ΔV satisfies Eq. (4.6) subject to boundary conditions (4.7)–(4.9), and ΔF satisfies Eq. (4.10), subject to boundary conditions (4.11)–(4.14). However, Eqs. (4.6) and (4.10) cannot be solved analytically in this case. Furthermore, it is difficult even to obtain numerical solutions; although these are parabolic partial differential equations, they are linked to each other through the two sets of boundary conditions. Fortunately, however, we can obtain approximate solutions as long as gt and ht are not too small.

If gt and ht are large, the investment and disinvestment boundaries will be far apart, and thus the two sets of boundary conditions will be relatively independent of each other. (Intuitively, if the investment boundary is hit, it is likely to take a long time before the disinvestment boundary is also hit, and vice versa.) In that case, the solutions to (4.6) and (4.10) will be of the time-separable form:

$$\Delta F = A(K)\,\theta^{\beta_1} e^{-gt} \tag{4.28}$$

and

$$\Delta V(K;\theta,t) = B(K)\theta^{\beta_2} e^{-ht} + \left(\frac{\eta-1}{\eta\delta}\right) K^{-1/\eta}\theta \tag{4.29}$$

with β_1, β_2, g, and h as given by Eqs. (4.18), (4.24), (4.21), and (4.27). The functions $A(K)$ and $B(K)$ and the critical values $\theta^*(K,t)$ and $\theta^{**}(K,t)$ can then be found from boundary conditions (4.8), (4.9), (4.12), and (4.13). Making these substitutions, the conditions become

$$B(K)(\theta^{**})^{\beta_2} e^{-ht} + \omega(K)\theta^{**} = A(K)(\theta^{**})^{\beta_1} e^{-gt} + k_0 e^{-st} \tag{4.30}$$

$$\beta_2 B(K)(\theta^{**})^{\beta_2-1} e^{-ht} + \omega(K) = \beta_1 A(K)(\theta^{**})^{\beta_1-1} e^{-gt} \tag{4.31}$$

$$A(K)(\theta^*)^{\beta_1} e^{-gt} = B(K)(\theta^*)^{\beta_2} e^{-ht} + \omega(K)\theta^* - k_0 e^{\rho t} \tag{4.32}$$

$$\beta_1 A(K)(\theta^*)^{\beta_1-1} e^{-gt} = \beta_2 B(K)(\theta^*)^{\beta_2-1} e^{-ht} + \omega(K) \tag{4.33}$$

For values of K and t, these four equations can be solved numerically for $A(K)$, $B(K)$, $\theta^*(K,t)$, and $\theta^{**}(K,t)$. We can also check the accuracy of these approximate solutions by determining whether $A(K)$ and $B(K)$ remain constant as t varies.

This is illustrated in Figure 4.6, which shows numerical solutions of Eqs. (4.30)–(4.33) for $A(K)$ and $B(K)$ for $K = 3$, as t varies from 0 to 9. (The parameter values are $r = \delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, $k_0 = 3.0$, and $\rho = s = 0.20$.) Observe that $A(K)$ and $B(K)$ become roughly constant once t is greater than about 2.

Figure 4.7 shows solutions for the investment and disinvestment thresholds, $\theta^*(K)$ and $\theta^{**}(K)$, as functions of K for $t = 2$ and 5. There are now three regions: If $\theta > \theta^*(K)$, the firm should invest immediately, increasing K (and thus increasing θ^*) until $\theta = \theta^*$. If $\theta < \theta^{**}(K)$, the firm should disinvest until $\theta = \theta^{**}(K)$. If $\theta^{**} \le \theta \le \theta^*$, the firm should take no action. Note that the thresholds $\theta^*(K)$ and $\theta^{**}(K)$ move apart over time, increasing the zone of inaction. This is illustrated in Figure 4.8, which shows θ^* and θ^{**} as functions of time for $K = 3$.
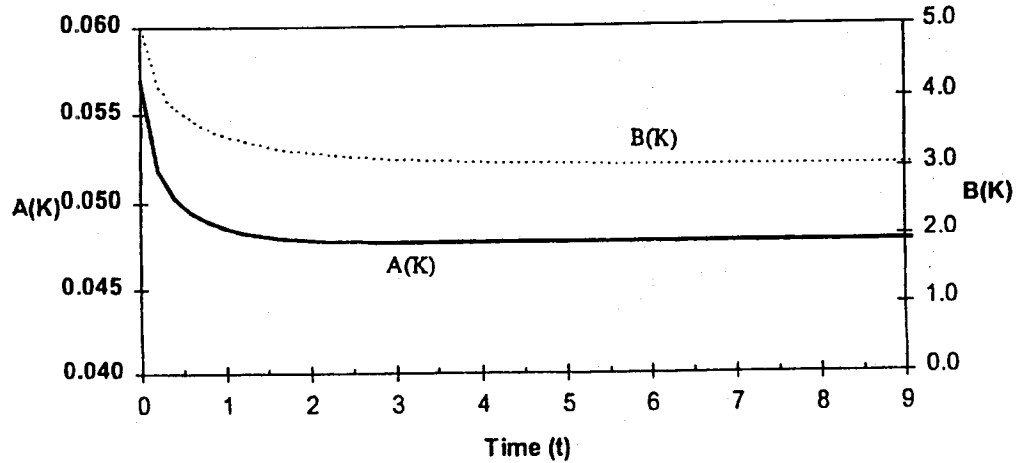
Figure 4.6   General case—numerical solutions for $A(K)$ and $B(K)$. Parameter values: $r = \delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, $k_0 = 3.0$, and $\rho = s = 0.20$.
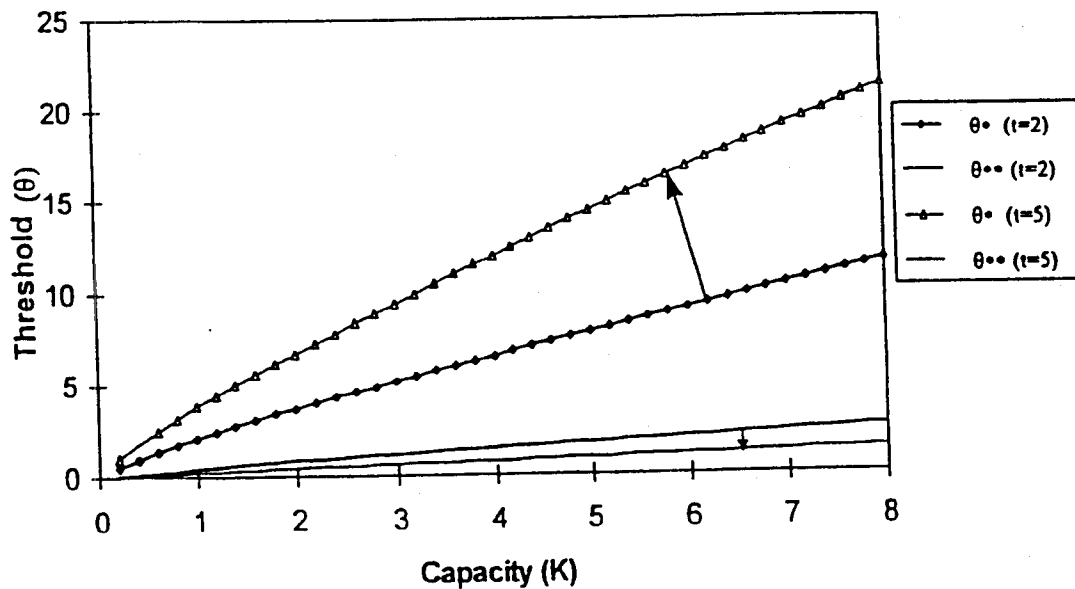

Figure 4.7   General case—investment and disinvestment thresholds, θ^* and θ^{**}, as functions of capacity K, for $t = 2$ and $t = 5$.

Figure 4.9 shows sample paths for the demand shift variable $\theta(t)$ and for capacity $K(t)$. Starting with no capital, the firm immediately invests at time t_0, bringing its capacity to K_0, such that $\theta_0 = \theta^*(K_0, t_0)$. From t_0 until t_1, $\theta^{**}(K_0, t) < \theta(t) < \theta^*(K_0, t)$, so the firm neither invests nor disinvests. Over this interval of time, the investment threshold θ^* increases gradually as the cost of adding capacity increases, while the threshold θ^{**} decreases gradually as the selling price of used capacity decreases. At time t_1, $\theta(t)$ hits the upper threshold θ^*, so the firm adds extra capacity. Over the interval t_1 to t_2, $\theta(t)$ is increasing, and capacity is increased from K_0 to K_1, so that $\theta^*(K, t) = \theta(t)$.[6] From t_2 to t_3, $\theta^{**}(K_1, t) < \theta(t) < \theta^*(K_1, t)$, so the firm is

---

[6] Note that the lower threshold $\theta^{**}(K, t)$ also increases as K increases.

again inactive. At time t_3, $\theta(t)$ hits the lower threshold θ^{**} and the firm disinvests. From t_3 to t_4, $\theta(t)$ continues to fall and the firm's capacity is gradually reduced from K_1 to K_2. After t_4 the firm is again inactive. Observe that as time goes on, $\theta^*(K_2, t)$ gradually increases and $\theta^{**}(K_2, t)$ decreases, so that the periods of investment or disinvestment become less and less frequent.
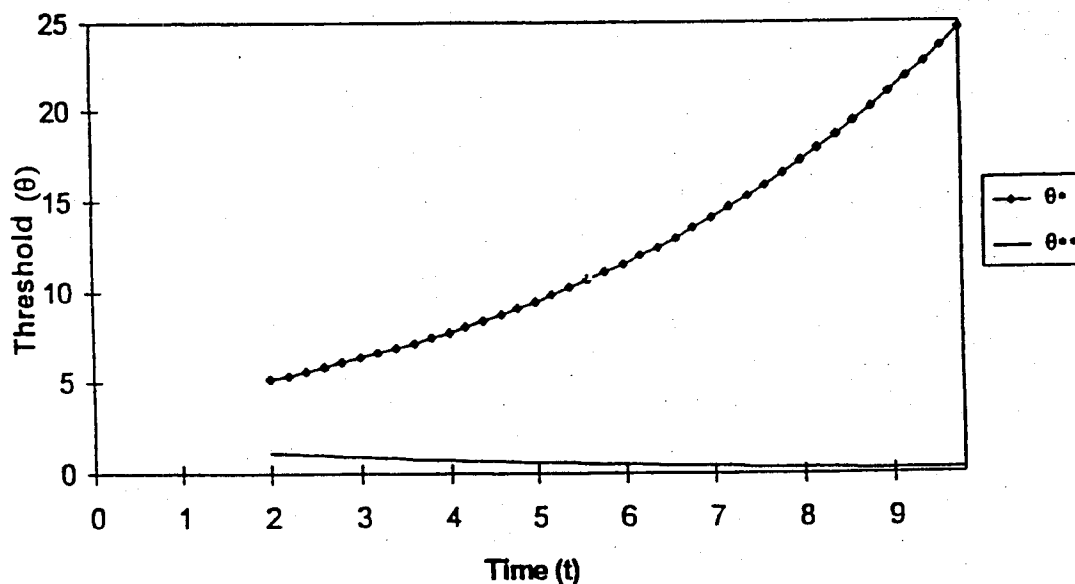

Figure 4.8 General case—movement of investment and disinvestment thresholds over time, for a given capacity $K = 3$.
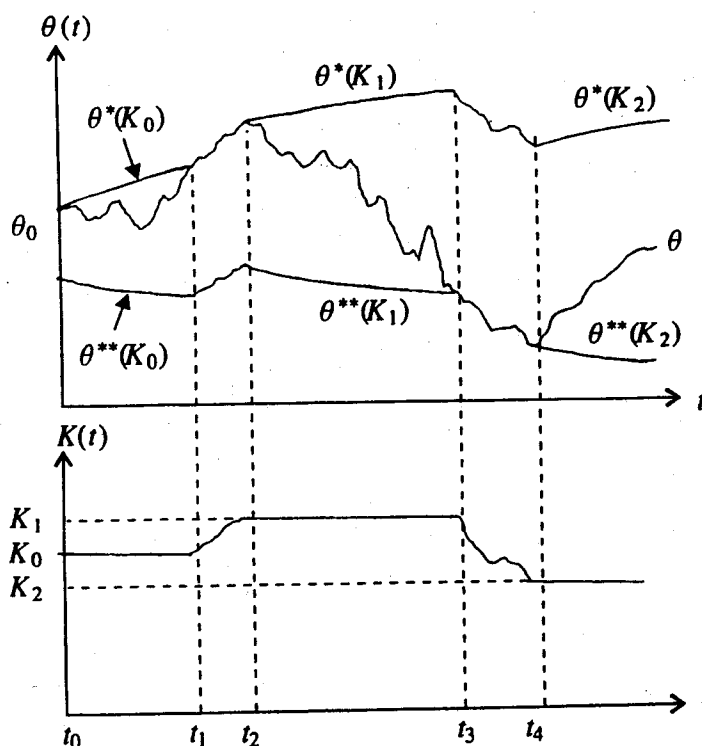

Figure 4.9 Optimal investment and disinvestment—sample paths of $\theta(t)$ and capacity $K(t)$.

4.4 STATIC VERSUS DYNAMIC EFFECTS OF SUNK COSTS

A significant difference between the *current* prices at which capital can be bought or sold will by itself create a zone of inaction in which the firm neither increases nor decreases capacity. This "static effect" of sunk costs of entry and exit is a standard result (e.g., see Chapter 7 of Dixit and Pindyck, 1996). Further, the expectation that the purchase and sales prices will diverge further *in the future* also affects the *current* investment thresholds. It is useful to separate these "static" and "dynamic" effects of limited expandability and reversibility.

Let us begin at some time t_1 when the purchase price of a unit of capital, k_p, exceeds the resale price, k_r. To determine the static effect of this differential, we calculate the investment and disinvestment thresholds, $\theta^*(K)$ and $\theta^{**}(K)$, under the assumption that these prices will remain fixed over time from t_1 onward.[7] Next, we calculate $\theta^*(K, t)$ and $\theta^{**}(K, t)$ under the assumption that at any future time $t > t_1$, the purchase price will be $k_p e^{\rho(t-t_1)}$ and the resale price will be $k_r e^{-s(t-t_1)}$. Although this "dynamic" $\theta^*(K, t)$ will rise over time and the "dynamic" $\theta^{**}(K, t)$ will fall over time, it is of interest to compare the static and dynamic thresholds at the initial time, t_1.

This is illustrated in Figure 4.10, starting out at $t = 5$ with $k_p = k_0 e$ and $k_r = k_0 e^{-1}$. (The other parameter values are $r = \delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, $k_0 = 3.0$, and $K = 3$.) Static thresholds are calculated assuming k_p and k_r remain fixed at these levels, while dynamic thresholds are calculated assuming that $k_p(t) = k_0 e^{\rho t}$ and $k_r(t) = k_0 e^{-st}$. Initially the zone of inaction is smaller in the dynamic case than in the static one. However, this zone of inaction grows in the dynamic case as θ^* rises and θ^{**} falls, and it eventually exceeds the zone in the static case.

Why is $\theta^*(K)$ initially lower in the dynamic case? There are two forces at work, with opposite effects. First, the fact that the purchase price of capital is expected to rise in the future reduces the value of the firm's call option on an incremental unit of capital, which reduces the value of waiting and so reduces $\theta^*(K)$. Second, the fact that the resale price of capital is expected to fall in the future reduces the value of the firm's put option to abandon installed capital, and so pushes $\theta^*(K)$ up. In the example shown in Figure 4.10 the first effect outweighs the second, so $\theta^*(K)$ falls.

The situation is similar with respect to $\theta^{**}(K)$. Again, the fact that the resale price k_r is expected to fall reduces the value of the firm's put option on an incremental unit of capital, which reduces the value of waiting to disinvest, and pushes $\theta^{**}(K)$ up. And the fact that the purchase price k_p is expected to rise reduces the value of the call option, which raises the cost of disinvesting now, and pushes $\theta^{**}(K)$ down. Once again, in this example the first effect outweighs the second, so $\theta^{**}(K)$ rises.

Of course, the magnitudes of these effects depend on various parameter values, besides those of ρ and s. For example, Figures 4.11 and 4.12 show the static and dynamic thresholds, and their movements over time, for two different values of the volatility of demand fluctuations, σ; when σ is larger, both the static and dynamic investment thresholds are higher, and the static and dynamic disinvestment thresholds are lower.

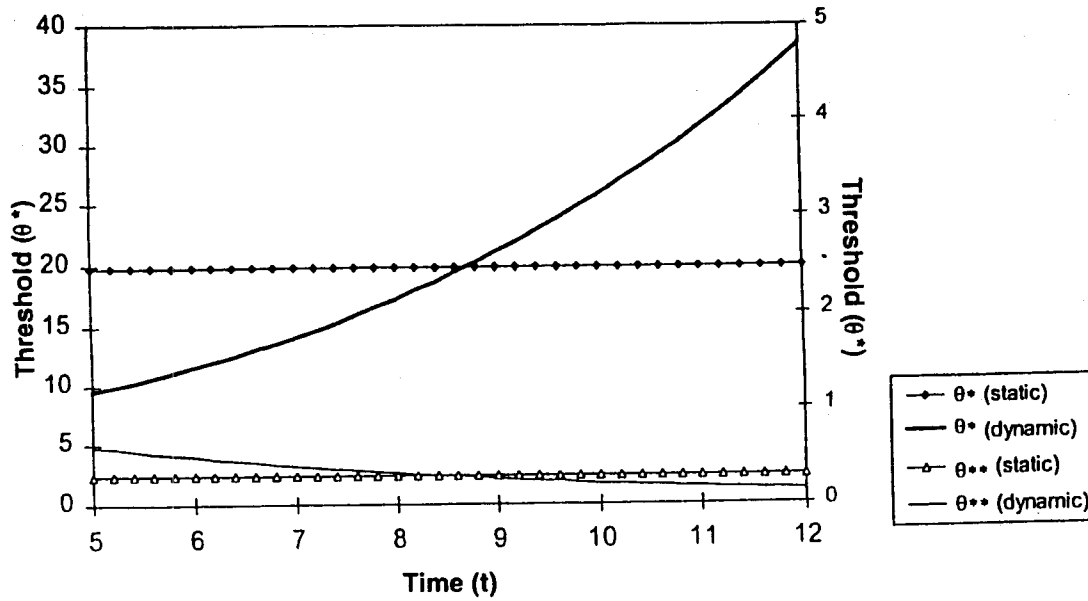[7] The thresholds will, of course, also be fixed through time.

Figure 4.10 Investment and disinvestment thresholds for static versus dynamic capital costs. Starting at $t = 5$, purchase price is $k_p = k_0 e$ and resale price is $k_r = k_0 e^{-1}$. Other parameters: $r = \delta = 0.05$, $\sigma = 0.40$, $\eta = 1.20$, $k_0 = 3.0$, $K = 3$, and $s = \rho = 0.2$.
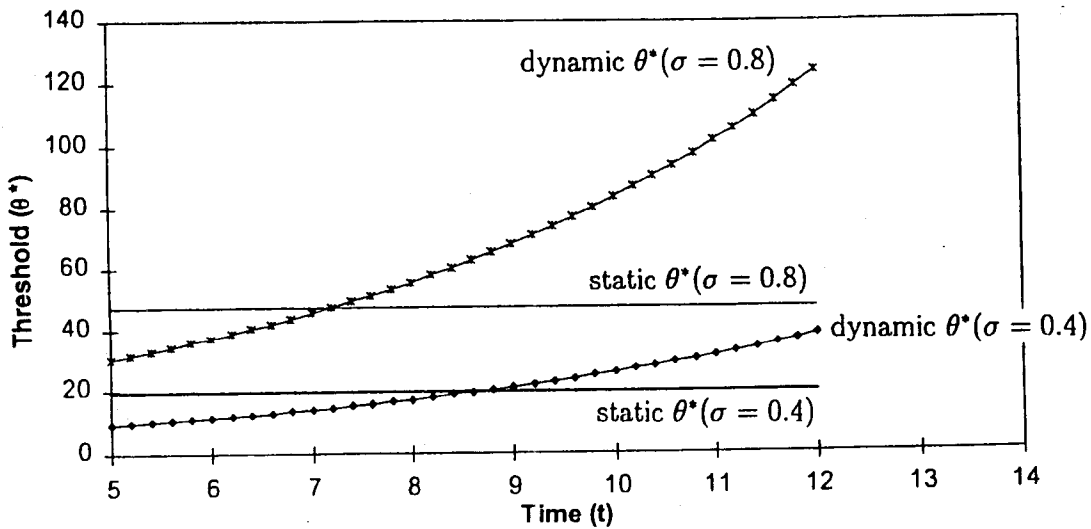
Figure 4.11 Static and dynamic investment thresholds for different levels of volatility ($\sigma = 0.4$ and 0.8).

Now that we have a better understanding of the source and nature of the dynamic effects, we can show their dependence on the rates of change of the purchase and resale prices of capital, ρ and s, respectively. We do this for a representative case in Table 4.1. The initial investment cost is 3 and the initial resale value is 1. This gap and the uncertainty ($\sigma = 0.4$) are so large that under static conditions ($\rho = s = 0$ from here on) the initial investment threshold is more than 6.5, and
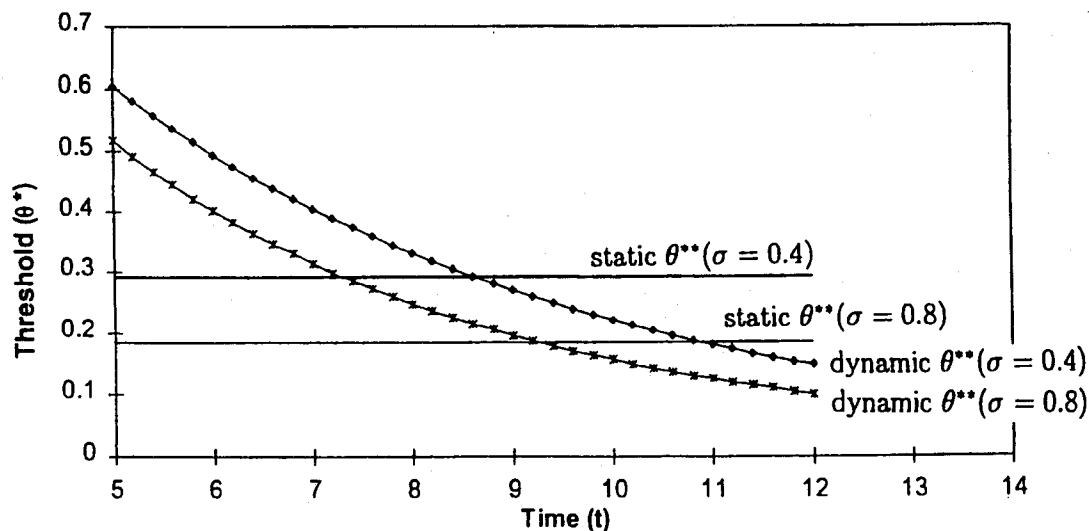
Figure 4.12   Static and dynamic disinvestment thresholds for different levels of volatility (σ = 0.4 and 0.8).

the initial disinvestment threshold less than 0.1. Table 4.1 shows what happens to the corresponding initial values of the dynamic thresholds as we vary ρ and s. Panel A shows that the dynamic investment threshold θ^* decreases as ρ increases, since the call option becomes less valuable. In fact, for very large values of ρ, it may become optimal to invest even though θ is less than the purchase price of capital; the latter is expected to grow so fast that it pays the firm to acquire the capital right away while it is cheap. Also, θ^* increases as s increases, because the put option of disinvesting is less valuable. These results confirm the intuition we gave above. But the numerical calculations reveal an interesting insensitivity: the effect of s on θ^* is very small. The gap between the two thresholds is sufficiently large that when θ is at the upper threshold, it is unlikely to fall to the lower threshold in the reasonable future. Therefore options that get exercised in that unlikely and remote eventuality do not have a significant effect on today's decision. The investment and disinvestment thresholds effectively become separated, as in the case discussed earlier in Sections 4.3.1 and 4.3.2. Given our assumption of a widening gap between purchase and resale prices, this seems to be a robust feature of our numerical calculations, and it may provide a useful simplification for the solution of combined investment and disinvestment problems when one or both of the degrees of irreversibility and uncertainty are large.[8]

Panel B of Table 4.1 shows the dynamic disinvestment threshold θ^{**} as we vary ρ and s. The results confirm the intuition stated above: the threshold rises as s increases because the put option exercised by disinvesting is less valuable, and it falls as ρ increases because the call option that would be acquired upon investment is less valuable. Again we find an effective separation of the two decisions: θ^{**} is relatively insensitive to changes in ρ, particularly for higher values.

[8] Bonomo (1994) discusses a similar issue in the context of models of impulse control, namely when a one-sided s–S rule is a sufficiently good approximation to a two-sided s–S policy.

Table 4.1

Effects of ρ and s on the Investment and Disinvestment Thresholds[a]

ρ	s			
	0.0	0.2	0.4	0.6
A. Initial dynamic investment threshold $\theta^*(t_1)$				
0.0	6.586	6.716	6.716	6.716
0.2	3.073	3.144	3.144	3.144
0.4	2.469	2.528	2.528	2.528
B. Initial dynamic disinvestment threshold $\theta^{**}(t_1)$				
0.0	0.098	0.246	0.283	0.300
0.2	0.084	0.201	0.231	0.244
0.4	0.084	0.200	0.230	0.244

[a] Parameters: $r = 0.05$, $\delta = 0.05$, $\sigma = 0.4$, $\eta = 1.2$, $k_p = 3$, $k_r = 1$. Static thresholds: $\theta^* = 6.58$, $\theta^{**} = 0.098$.

4.5 CAPACITY-DEPENDENT COSTS

In this section we examine endogenous variations in the costs of investing and disinvesting. We briefly consider situations in which the ability of the firm to add or reduce capacity in the future is dependent on its own past actions, in particular on the amount of capacity that it already has in place, rather than the amount of time that has elapsed.

Specifically, we assume that the firm can add capacity at any time in the future at a cost $k_0 + \rho(K)$ per unit of capital, with $d\rho/dK > 0$. In effect, it becomes more expensive to add capacity the more capacity the firm already has. Such limits to expandability may be driven by market parameters, such as population and available land, but depend on the firm's own actions rather than the actions of its competitors. In the simplest case, the incremental cost of capacity expansion may be linear, i.e., $\rho(K) = \rho K$. Then, if $\rho = \infty$ there is no expandability (even from zero), while if $\rho = 0$ there is complete expandability.

Likewise, we assume that the firm can sell off capacity at any time in the future such that if its current capacity is K, it will receive $s_0 k_0 + s_1 \rho(K)$ for an incremental unit, with $0 \leq s_0 \leq 1$ and $s_1 < 1$. In the simplest linear case, the firm receives $s_0 k_0 + s_1 \rho K$. Thus the degree of irreversibility is different for each marginal unit. We would expect that irreversibility would be greater the greater is K, since the demand for used industry-specific capital is likely to be smaller the greater is the amount of installed capacity already in place. In this case, $s_1 < s_0$. Also, it might be the case that $s_1 < 0$, so that if K is large enough the firm receives a negative amount on its sale of an incremental unit. This could occur, for example, if the firm faces land reclamation costs. Finally, if $s_0 = s_1 = 0$ there is complete irreversibility, and if $s_0 = s_1 = 1$ there is complete reversibility.

Equations (4.6) and (4.10) will again apply for ΔV and ΔF, but now without the ΔV_t and ΔF_t terms. Hence the investment problem is now much simpler: θ^{**} and θ^* each depends only on K, and not on t. In the general case, the boundary conditions again result in four nonlinear equations for $A(K)$, $B(K)$, $\theta^{**}(K)$, and $\theta^*(K)$, which can easily be solved numerically for each value of K.

We do not present numerical solutions here because the basic effects are similar to those in the entry/exit models discussed in Dixit and Pindyck (1996). What is different here is that the investment and disinvestment thresholds, $\theta^*(K)$ and $\theta^{**}(K)$, depend on K. As K increases, the direct value (i.e., the present value of the marginal revenue stream) of an incremental unit of capital falls, as does the value of the call option on the unit. The drop in the direct value raises the investment threshold $\theta^*(K)$, while the drop in the option value reduces the threshold. The first effect dominates, however, so that $\theta^*(K)$ increases with K. The opposite may be true for the disinvestment threshold, $\theta^{**}(K)$. As K increases, the direct value of the incremental unit of capital falls, but the value of the put option on the unit increases. The degree of reversibility (i.e., the resale value of capital) determines whether the value of the put option exceeds the direct value of the incremental unit. If it does, $\theta^{**}(K)$ will also increase with K.

4.6 CONCLUDING REMARKS

We have analyzed how the call and put options associated with limited expandability and reversibility interact to affect a firm's optimal capacity decisions and the evolution of capacity over time. Expandability and reversibility can take a variety of forms. For example, a firm might be able to expand only at specific points in time (e.g., a forest products or extractive resource firm might have to wait for the government to auction off land or resource reserves), in which case its ability to sell existing capital might occur unpredictably as a Poisson arrival (e.g., when there are very few potential buyers who might become interested in a specialized piece of capital). We have examined here only a very special form of expandability and reversibility—namely, one in which capital purchase and sales prices evolve exogenously with time, or endogenously with the level of installed capacity. Nonetheless, we believe this analysis helps to elucidate the basic effects. Most importantly, we can see how the future rates of growth of the investment cost and the resale price of capital affect the values of the call and put options associated with expansion and disinvestment. Our numerical results reveal that the investment and disinvestment decisions become separated when the initial gap between the purchase and resale prices of capital is substantial, or the rate at which this gap grows is rapid. This condition simplifies both the analytical and numerical solution of these problems.

This analysis can be extended or generalized in several ways. (1) Various aspects of increased realism can be added to the model, albeit at the cost of increased complexity. For example, we took variable costs to be zero, so that we could ignore the firm's operating options. It would be messy but not overly difficult to extend our model by including a positive variable cost and allowing the firm to vary its capacity utilization. Similarly, other operating options can be added. (2) We considered a firm's decision problem in isolation, treating as exogenous the rates of variation

of the purchase and resale prices, whether as functions of time or as functions of existing capacity. These can be endogenized in a more complete general equilibrium analysis. (3) We confined ourselves to a setting in which the maximization problem is well behaved. An aspect of this was our assumption in the previous section that the purchase price of capital increases with capacity. If instead we allowed the purchase price to decrease with capacity, the firm would enjoy increasing returns to capacity expansion and its optimal policy could consist of infrequent large jumps in its capital stock. Dixit (1995) shows how to find the optimal timing and size of such jumps, but numerical work for more specific parameterized models can provide further useful insights. (4) The physical depreciation of capital can be modeled more realistically. The resale price of a newly installed machine would equal its purchase price, but would fall with the *age* of the machine, not with *calendar time* as in our present model. This, however, would require keeping track of the installation dates or age profiles of the entire stock of machines, making the state variable infinite dimensional, and presenting daunting modeling and numerical solution challenges.

ACKNOWLEDGMENTS

REFERENCES

Abel, A. B., A. K. Dixit, J. C. Eberly, and R. S. Pindyck. (1996). Options, the Value of Capital, and Investment. *Quarterly Journal of Economics* 111(446), 753–777.

Bonomo, M. (1994). Optimal Two-Sided and Suboptimal One-Sided Pricing Rules. Working Paper No. 313, Pontificia Universidade Catolica, Rio de Janeiro.

Brennan, M. J., and E. S. Schwartz. (1985). Evaluating Natural Resource Investments. *Journal of Business* 58, 135–157.

Dixit, A. K. (1989). Entry and Exit Decisions Under Uncertainty. *Journal of Political Economy* 97(3), 620–638.

Dixit, A. K. (1992). Investment and Hysteresis. *Journal of Economic Perspectives* 6(1), 107–132.

Dixit, A. K. (1995). Irreversible Investment with Uncertainty and Scale Economies. *Journal of Economic Dynamics and Control* 19(1), 327–350.

Dixit, A. K., and R. S. Pindyck. (1996). *Investment Under Uncertainty*. Princeton, NJ: Princeton University Press (second printing).

Kulatilaka, N. (1995). Operating Flexibilities in Capital Budgeting: Substitutability and Complementarity in Real Options. In L. Trigeorgis (ed.), *Real Options in Capital Investment*. New York: Praeger.

McDonald, R., and D. R. Siegel. (1985). Investment and the Valuation of Firms When There Is an Option to Shut Down. *International Economic Review* 26, 331–349.

Myers, S. C., and S. Majd. (1984). Calculating Abandonment Value Using Option Pricing Theory. Working Paper No. 1462-83, MIT Sloan School of Management.

Pindyck, R. S. (1988). Irreversible Investment, Capacity Choice, and the Value of the Firm. *American Economic Review* 79(5), 969–985.

Pindyck, R. S. (1991). Irreversibility, Uncertainty, and Investment. *Journal of Economic Literature* 29(3), 1110–1152.

Trigeorgis, L. (1993). The Nature of Option Interactions and the Valuation of Investments with Multiple Real Options. *Journal of Financial and Quantitative Analysis* 28(1), 1–20.

Trigeorgis, L. (1996). *Real Options*. Cambridge, MA: MIT Press.