MONOPOLY WITH TIMING AND SCALING OPTIONS

Sigbjørn Sødal*

Agder University College, Norway

(December 2001)

Abstract: We develop a model of the investment behavior of a firm that faces a stochastic, downward-sloping demand curve. The firm's challenge is to determine the optimal scale and time of an investment, so there is a potential for market power in the sense of markup pricing along two dimensions: static market power along a quantity dimension, and dynamic market power along a time dimension. Depending on the specific assumptions, either dimension will be more or less relevant. For example, the option to wait may be useless if the uncertainty of demand is low and the demand curve is not very elastic. Then the decision of the firm simplifies to that of a standard monopoly model. In other cases, the option to wait prevails. Typically, the latter happens when there is much uncertainty and the demand curve is fairly elastic. The formal model is illustrated by decisions in the real estate industry.

Keywords: irreversible investment, demand uncertainty, market power, markup pricing.
JEL classification: D92, D81, D42.

1. Introduction

A monopolistic firm can be seen as the owner of one or more options that are exercised by exploiting market power in the sense of pricing above marginal costs. Two such options are of special interest. The standard monopoly model – known from any introductory course in microeconomics – is based on the option to reduce production to increase the price above the marginal production cost. The standard real options model of irreversible investment under price uncertainty – see Dixit and Pindyck (1994) – focuses on the option to postpone an investment until the revenue that is obtained exceeds the investment cost. In both models, the optimal decision consists of maximizing the difference between revenue and cost.

Since the scale of production often results from an irreversible investment, these two options ought to be discussed in relation. We should aim at a description of monopoly where the firm is allowed to delay as well as to scale production, subject to demand conditions that may change over time.

---

* Agder University College, Faculty of Economics and Social Sciences, Serviceboks 422, 4604 Kristiansand, Norway. Email: Sigbjorn.Sodal@hia.no. I am grateful to Geir B. Asheim for helpful comments.

For example, consider a firm in the real estate industry that has acquired a certain piece of land. The firm is to find the optimal time and the optimal number of homes to build. For technological, economic or political reasons, a unique lot size is needed, and all homes must be built in one batch. Most likely, the price that can be achieved will be an increasing function of the lot size. This leads to a downward-sloping demand curve. By choosing a small lot size, many homes can be sold, but the price will be low; by choosing a larger lot size, fewer homes can be sold, but the price will be higher. In addition, the location of the demand curve could be uncertain, as aggregate demand for such homes typically fluctuates over time. Therefore the firm must also consider postponing the investment to obtain higher prices. Hence, the optimal decision consists of two elements: when and how much to invest.

The objective of this paper is to discuss a formal model based on these ideas. The model is general, but it is convenient to illustrate with the real estate industry when explaining its various assumptions. Below we describe the two standard models upon which the model is built.

First, suppose a monopolist with a simple cost function – a fixed cost and a constant marginal cost – is facing a downward-sloping demand curve. The profit function to be maximized is

(1) $$w = q(p)(p - c) - a,$$

where p is the price, q(p) is the demand function, a is the fixed cost, and c is the marginal cost. Under certain well-known conditions, the optimal price is given by the markup

(2) $$\frac{p - c}{p} = \frac{1}{e},$$

where e=–(dq/dp)/(q/p) is the magnitude of the price elasticity of demand, and e>1 is assumed.

Next, consider the standard investment model by McDonald and Siegel (1986), answering the following question: When is it optimal to invest C to obtain P_t, where C is fixed, while P_t fluctuates according to a general Ito process

(3) $$dP_t = f(P_t)dt + g(P_t)dz_t,$$

assuming a constant discount rate ρ? Here dz_t is the Wiener process, and $f(P_t)$ and $g(P_t)$ are continuous functions. For comparison and reference, let us assume that the stochastic process takes off right after a fixed acquirement cost A has been paid. The initial value of the process, P_0, is low in a sense to be explained later.

The solution to this dynamic problem can also be described in terms of a markup, as the optimal decision is to invest as soon as P_t reaches a fixed value higher than C. Following Dixit et al. (1999), the expected net present value of investing at an arbitrary $P \geq P_0$, evaluated at time zero, can be written as

(4) $$W = Q(P)(P - C) - A.$$

Here $Q(P) \equiv E[e^{-\rho T}]$ is an alternative expression for the expected discount factor when the process is to move from P_0 to P for the first time, so T is the first hitting time. Eq. (4) can be explained as follows: The acquirement cost A applies at time zero, and is subtracted. The next investment brings the revenue P and the cost C. Since this occurs in the future, the net benefit, (P−C), must be discounted. The expected discount factor, Q(P), which depends on the process (3), is strictly decreasing in P because it takes longer time to reach P the farther it is from P_0. Hence Q(P) is analogous to a demand function, P is analogous to a price, and A and C are analogous to a fixed cost and a constant marginal cost, respectively. The optimal markup therefore becomes

(5) $$\frac{P - C}{P} = \frac{1}{\varepsilon},$$

where $\varepsilon = -(\partial Q/\partial P)/(Q/P)$ is the magnitude of the elasticity of the discount factor.

Eq. (2) and Eq. (5) show that the two models coincide in a technical sense even if the assumptions are very different: The option to wait is ignored in the static model, while the option to scale production is ignored in the dynamic model. The general model that follows relaxes both constraints and reveals some economic implications of the analogy. Note that the net profit margin must be high enough to cover the fixed cost in both models above, as well as in our more general model; i.e., $W \geq 0$ is required. The dynamic setting also requires $Q \leq 1$. Combined with Eq. (4) and Eq. (5), this implies $A(\varepsilon - 1) \leq C$.

2. The model

Suppose a firm has acquired the right to produce a specific good. Referring to the real estate industry, we may imagine that the firm owns a piece of land. Demand at time t is given by

(6) $$q_t = q(x_t, p_t),$$

where $q(x_t,p_t)$ is increasing in x_t and decreasing in p_t. The variable x_t determines the stochastic development of demand, and is given by a general Ito process similar to Eq. (3):

(7) $$dx_t = f(x_t)dt + g(x_t)dz_t.$$

Referring to the example again, q_t may represent the number of homes. Uncertain market conditions imposed by the stochastic process x_t cause the relation between price and quantity to change over time. For now, we assume that the initial value x_0 is so low that immediate production – i.e., building homes on the land – is not optimal.

The investment cost C(q) applies to production, where dC/dq>0. There is no disinvestment opportunity, so even if C might be a net present value, we do not consider how this cost is distributed over time. We also exclude possible new investments to expand capacity. (This is a reasonable assumption in the real estate industry, as it is generally not allowed to split up lots in an existing residential area.)

The decision problem above boils down to the question of when to make a single investment, and what scale to choose. Thus, the firm is not a durable monopoly competing with itself as in Coase (1972). More generally, output will not necessarily be socially optimal as in rational expectations models without market distortions; see Lucas and Prescott (1971).

The revenue by investing at scale q and time t equals $p_t(x_t,q)q$, where $p_t(x_t,q)$ is the inverse demand function implied by Eq. (6). For any fixed q, it follows from McDonald and Siegel (1986) that the optimal time to invest will be as soon as the stochastic variable x_t reaches a certain value higher than x_0. Since this holds for any scale of production, it also holds for the optimal scale. By similar arguments as in the introduction, it follows that the expected net present value of investing at an arbitrary $x \geq x_0$, choosing scale q, can be written as follows:

(8) $$W(x,q) = Q(x)\big(R(x,q) - C(q)\big) - A.$$

Here $R(x,q) \equiv p(x,q)q$ is the revenue from sale, while $Q(x) \equiv E[e^{-\rho T}]$ is the expected discount factor as already described. The profit function of Eq. (8) generalizes Eq. (4) and can be explained in similar ways. This is the core of the model, since the optimal policy is found simply by maximizing W with respect to x and q. Assuming that the function is smooth, the two first-order conditions for a maximum can be expressed by a set of elasticities:

(9a) $$\frac{R-C}{R} = \frac{\varepsilon_{Rx}}{\varepsilon_{Qx}}$$ (9b) $$\frac{R}{C} = \frac{\varepsilon_{Cq}}{\varepsilon_{Rq}} .$$

The elasticities are defined by $\varepsilon_{Rx}=(\partial R/\partial x)/(R/x)$, $\varepsilon_{Qx}=-(\partial Q/\partial x)/(Q/x)$, $\varepsilon_{Rq}=(\partial R/\partial q)/(R/q)$, and $\varepsilon_{Cq}=(\partial C/\partial q)/(C/q)$. Eq. (9a) shows the optimal revenue as a markup over the investment cost. The markup is high if ε_{Qx} is small – i.e., if the discount factor function is not very elastic. Typically, this is the case if the process is highly stochastic or the discount rate is low. The elasticity ε_{Rx} may pull in the other direction, since the revenue may increase less than proportionally in x.

Eq. (9b) reflects the influence of marginal costs. Since $R>C$ is required for positive profit, $\varepsilon_{Rq}<\varepsilon_{Cq}$ must hold. In practice, this implies restrictions with respect to economies of scale.

The acquirement cost A is irreversible, so it does not affect the decision on when to undertake the next investment. The marginal valuation of Eq. (8) is to maximize $Q(x)(R(x,q)-C(q))$ regardless of the exact initial value x_0. At the optimal time, we have $x_0=x$, where x is fixed, and the expression to be maximized simplifies to $R(x,q)-C(q)$. Hence, the decision reduces to that of static monopoly as soon as the value of waiting has been fully exploited.

In a free market, one can expect that the price of land is bid up until W=0. This implies $A=Q(R-C)$, where Q is the optimal discount factor evaluated at time zero, and R and C also carry optimal values.

The next section exemplifies this model using two common stochastic processes: geometric and arithmetic Brownian motions. Both processes are combined with isoelastic demand functions and simple cost functions.

3. Examples

Geometric Brownian motion

Suppose demand at time t is given by the equation

(10) $q_t = (x_t / p_t)^s$,

where p_t is the price, s>1 is a constant price elasticity of demand, and x_t is a geometric Brownian motion with drift μ (<ρ) and volatility σ (≥0):

(11) $dx_t = \mu x_t dt + \sigma x_t dz_t$.

As shown by Dixit et al. (1999), the discount factor associated with this process is

(12) $Q(x_0, x) = (x_0 / x)^\beta$,

where β (>1) is the positive solution to the following quadratic equation in m:

(13) $\frac{1}{2}\sigma^2 m(m-1) + \mu m - \rho = 0$.

Eq. (12) implies that the discount factor of the geometric Brownian motion is analogous to a demand function with a constant elasticity, β.

The production cost is assumed to be

(14) $C(q) = \begin{cases} cq, & q \le 1 \\ \infty, & q > 1 \end{cases}$

where c>0 is constant. Referring to the real estate industry, the cost of each home is constant. The capacity constraint, which for simplicity is scaled to unity, establishes sufficiently decreasing returns to scale. For example, the constraint may follow from a minimum lot size determined by law.

Eq. (10) implies that the revenue from investing at scale q as soon as an arbitrary $x>x_0$ has been reached, equals $R(x,q)=xq^{(s-1)/s}$. Then the expected net present value of the firm (at time zero) can be written as

(15) $$W(x,q)=\left(\frac{x_0}{x}\right)^{\beta}\left(q^{(s-1)/s}x-cq\right)-A,$$

where $0\leq q\leq 1$ is required. The elasticities of Eq. (9) are $\varepsilon_{Rx}=1$, $\varepsilon_{Qx}=\beta$, $\varepsilon_{Cq}=1$, and $\varepsilon_{Rq}=(s-1)/s$. Eq. (9a) implies $(R-C)/R=1/\beta$, while Eq. (9b) can be shown to yield $(R-C)/R=1/s$. Both equations cannot be correct unless s and β happen to coincide, so in this case we end up in a corner solution. It can be shown that W decreases (increases) strictly in x along the curve $\partial W/\partial q=0$ when s is smaller (larger) than β. Likewise, W increases (decreases) strictly in q along the curve $\partial W/\partial x=0$ when β is smaller (larger) than s. Combining these results, it follows that the exact corner solution depends on the sign of $\beta-s$.

The dynamic component of the model is too weak when $s<\beta$, in the sense that the gain from waiting for higher demand never exceeds the cost. Firm value is maximized in such cases by investing immediately (when $x=x_0$). By Eq. (10), this yields $W=(x_0/p)^s(p-c)-A$, so the optimal decision simplifies to that of a static monopoly model with constant price elasticity of demand. As long as x_0 is low enough, the optimal price-cost ratio becomes $p/c=s/(s-1)$. The capacity constraint determines the outcome in the opposite case. The solution is plotted in Fig. 1 for two levels of x_0. The constraint binds when $x_0=x_0^{high}$ as the marginal revenue (MR^{high}) curve intersects with the marginal cost (MC) curve at the vertical part of the latter curve. The constraint does not bind when $x_0=x_0^{low}$.

The dynamic component of the model dominates when $\beta<s$. In such cases the maximum scale of production (q=1) should be chosen, and Eq. (15) simplifies to Eq. (4) except for notation. This implies that the exact price elasticity of demand, s, is irrelevant as long as it exceeds β. Market power is exploited exclusively along the time dimension as in the standard investment model. Eq. (5) determines the optimal price-cost ratio, $p/c=\beta/(\beta-1)$. The solution is plotted in Fig. 2. Note that the marginal revenue (MR) curve intersects with the marginal cost (MC) curve at the vertical part of the latter curve because $\beta<s$.

It can be shown that $\partial\beta/\partial\mu<0$, $\partial\beta/\partial\sigma<0$, and $\partial\beta/\partial\rho>0$, so for a fixed price elasticity of demand, s, the dynamic option prevails if the drift or uncertainty of demand is high enough or the discount rate is low enough. Likewise, for fixed dynamics as reflected in a constant β, the static option prevails if demand is not very elastic.

These results can be explained by the specification of demand. It follows from Eq. (10) that the price process for a fixed q is geometric Brownian with drift μ and volatility σ. As argued by Dixit et al. (1999, p. 184), a geometric Brownian process implies an isoelastic discount factor because the probability distribution for percentage changes in the stochastic variable is independent of the current value. Just as the percentage change of demand following a percentage price change is constant in a static monopoly model with isoelastic demand, the change in the discount factor caused by a percentage price change in the dynamic model is independent of the current price. Holding such a discount factor up against an isoelastic demand function and a strict capacity constraint, while allowing for market power in both dimension, one dimension must dominate. This will be the dimension with the better opportunity to exploit market power – i.e., where the elasticity is lower.

Note also the potential differences with respect to deadweight loss. There is a deadweight loss as in a static monopoly model with constant marginal costs when $s<\beta$ and initial demand is sufficiently low (Fig. 1). When $s>\beta$ there is no deadweight loss (Fig. 2).

Arithmetic Brownian motion

Suppose demand is still given by Eq. (10) except that x_t is arithmetic Brownian:

(16) $$dx_t = \mu dt + \sigma dz_t .$$

As shown by Dixit et al. (1999), the discount factor associated with this process is

(17) $$Q(x_0, x) = e^{-\alpha(x-x_0)} ,$$

where α is the positive solution to the quadratic equation

(18) $$\tfrac{1}{2}\sigma^2 m^2 + \mu m - \rho = 0 .$$

Eq. (17) shows that the discount factor of the arithmetic Brownian motion is analogous to an exponential demand function.

The capacity constraint of the previous example is no longer needed, so now we simply define the production cost by $C(q)=cq$, where $c>0$ is a constant. The expected net present value of investing at scale q when an arbitrary $x>x_0$ has been reached, becomes:

(19) $$W(q,x) = e^{-\alpha(x-x_0)}\left(q^{(s-1)/s}x - cq\right) - A.$$

The elasticities of Eq. (9ab) are $\varepsilon_{Qx}=\alpha x$, $\varepsilon_{Rx}=1$, $\varepsilon_{Cq}=1$, and $\varepsilon_{rq}=(s-1)/s$, implying the following optimal policy:

(20a) $$x = \frac{s}{\alpha}$$ (20b) $$q = \left(\frac{s-1}{\alpha c}\right)^{s}$$

This solution is well defined as long as $x_0<s/\alpha$. In the opposite case, production should take place immediately, and the scale of production should be chosen as in static monopoly.

For fixed dynamics as reflected in a constant α, Eq. (20a) implies that more waiting will take place by increasing s. In other words, the option to wait becomes more important when the demand function gets more elastic. The variable α mimics β in underlying parameters, so we have $\partial\alpha/\partial\mu<0$, $\partial\alpha/\partial\sigma<0$, and $\partial\alpha/\partial\rho>0$. Thus, higher drift and uncertainty of demand also lead to more waiting, while a higher discount rate leads to less waiting.

Eq. (20b) shows that the optimal scale of production decreases along with s. Referring to the real estate industry again: lot sizes larger than minimum becomes more likely by making demand less elastic. The lot size also increases (q decreases) by increasing c or decreasing α.

These results look reasonable in a broader context. For example, we usually see expensive homes placed on large lots. Moreover, the typical lot size may decrease due to scarcity of land when demand is expected to increase (reflected in a higher μ). In such times the equilibrium price of land increases – i.e., the fixed cost A corresponding to W=0.

By inserting Eq. (20a,b) into Eq. (10), the optimal price-cost ratio becomes $p/c=s/(s-1)$, so the price coincides with that of static monopoly under constant marginal costs. This confirms that the model simplifies to static monopoly when the value of waiting has been fully exploited.

4. Concluding remarks

The model of this paper derives the optimal investment policy for a monopolistic firm that is to find the time and scale of an investment subject to a stochastic, downward-sloping demand curve. The solution depends on what market power dimension is more important. Static market power along the quantity dimension tends to dominate when the price elasticity of demand is low. Dynamic market power along the time dimension tends to dominate when the uncertainty or drift of demand is high or the discount rate is low.

In case of isoelastic demand and a geometric Brownian demand process as in our first example, the two options are summed up by two constant elasticities (s and β), but one option prevails. The exact value of the higher elasticity is irrelevant as long as the marginal cost is constant up to a fixed capacity. In other cases of interest – here exemplified by a constant-elasticity demand curve and an arithmetic Brownian process – typical results are not as extreme. Then both options may be valuable, but the general characteristics remain. High uncertainty or drift of demand, and low discount rates, tend to increase the value of options to wait, while less elastic demand tends to increase the value of options to scale production.

As far as economic modeling is concerned, it follows from the first example, in particular, that it can make sense to ignore the time dimension as in the standard monopoly model when the market is fairly stable, and demand is not very price elastic. When demand is relatively elastic while the market is volatile, one may rather fix the scale of production, and focus on value from waiting as in the standard investment model. Even if this means that both standard models may be reasonable simplifications under many circumstances, the empirical evidence indicates that combined use of the two options to exploit market power also can be useful. According to Dixit and Pindyck (1994, p. 136), dynamic markups exceeding 100 percent may well occur. This implies elasticities of the discount factor somewhere between 1 and 2. Typical price elasticities of demand are in the same range in many monopolized markets.

References

Coase, R. H., 1972. Durability and Monopoly. *Journal of Law and Economics* 15, 143-149.

Dixit, A. K., Pindyck, R. S., 1994. *Investment Under Uncertainty*. Princeton University Press.

Dixit, A., Pindyck, R. S., Sødal, S., 1999. A Markup Interpretation of Optimal Investment Rules. *Economic Journal* 109(455), 179-189.

McDonald, R., Siegel, D., 1986. The Value of Waiting to Invest. *Quarterly Journal of Economics* 101, 707-728.

Lucas, R. E., Prescott, E. C., 1971. Investment under Uncertainty. *Econometrica* 39, 659-681.

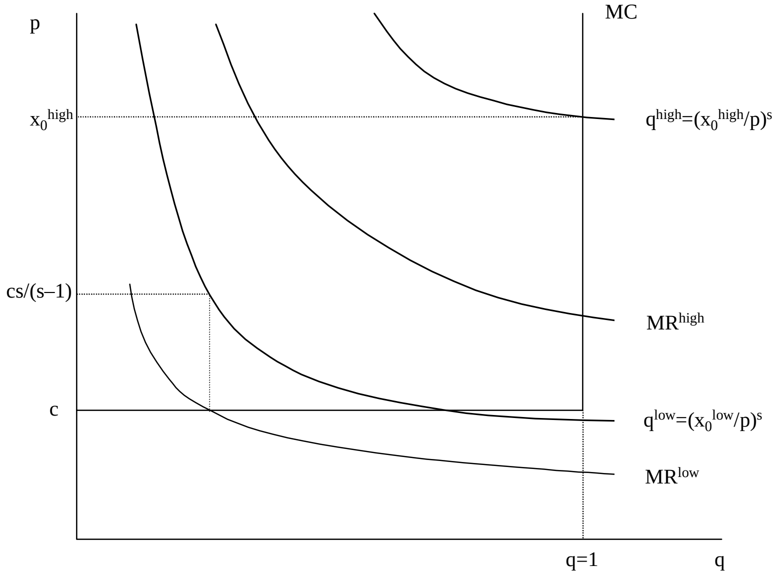

Fig. 1. Optimal investment without value of waiting (s<β).

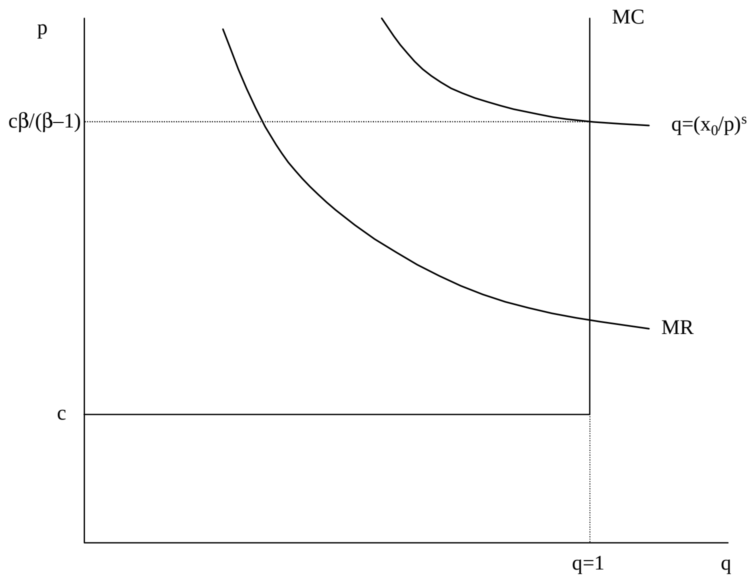

Fig. 2. Optimal investment with value of waiting (s>β).