

SUNK COSTS AND RISK-BASED BARRIERS TO ENTRY*

by

Robert S. Pindyck
Massachusetts Institute of Technology
Cambridge, MA 02142

This draft: February 13, 2009

Abstract: In merger analysis and other antitrust settings, risk is often cited as a potential barrier to entry. But there is little consensus as to the kinds of risk that matter — systematic versus non-systematic and industry-wide versus firm-specific — and the mechanisms through which they affect entry. I show how and to what extent different kinds of risk magnify the deterrent effect of exogenous sunk costs of entry, and thereby affect industry dynamics, concentration, and equilibrium market prices. To do this, I develop a measure of the “full,” i.e., risk-adjusted, sunk cost of entry. I show that for reasonable parameter values, the full sunk cost is far larger than the direct measure of sunk cost typically used to analyze markets.

JEL Classification Numbers: L40; L10, D43

Keywords: Entry barriers, sunk costs, investment decisions, risk, market power, antitrust.

*My thanks to Dennis Carlton, Wouter Dessein, Joe Farrell, Michael Riordan, Manuel Trajtenberg, Yoram Weiss, and seminar participants at Brandeis, Columbia, and Tel Aviv Universities for helpful comments and suggestions, and to Eric Wang for superb research assistance.

1 Introduction.

Barriers to entry are a fundamental determinant of market structure, and play a central role in merger analysis and other antitrust settings. When evaluating a proposed merger, for example, officials at the Department of Justice or Federal Trade Commission estimate the merger’s likely impact on prices, and the extent to which that impact would be limited by the potential entry of new firms.

Barriers to entry can arise from a variety of sources, but if entry requires large sunk costs, the risks associated with post-entry profits can be particularly important. This is especially true for rapidly evolving industries and markets for new products. But there is little consensus as to the kinds of risk that should matter and the mechanisms through which they affect entry and industry equilibrium. Indeed, the term “risk” is often used loosely in antitrust settings.¹

I use the term “entry barrier” to refer to any additional cost an entrant must pay that has already been paid by an incumbent, sufficient to allow the incumbent to raise price without inducing entry.² An entry barrier can be equivalently defined in terms of its effects — it limits the number of firms in the industry and increases price-cost margins. Thus large sunk costs are clearly an entry barrier; by creating scale economies, they lead to an industry equilibrium with relatively few firms. This is the case whether or not incumbent firms or potential entrants face any risk with respect to future cash flows. The question I address is how and to what extent different types of risk magnify these effects of sunk costs.

I treat risk as a basic structural feature of a market, and examine how risk and sunk costs interact to create entry barriers and affect industry concentration and market price. I

¹For example, in its merger policy guidelines, the European Commission (2004) discusses potential entry as a constraint on the behavior of post-merger incumbents, and stresses the importance of risk: “Furthermore, high risks and costs of failed entry may make entry less likely.” (Para. 69.) “Potential entrants may encounter barriers to entry which determine entry risks and costs and thus have an impact on the profitability of entry.” (Para. 70.) Likewise, the DOJ and FTC’s *Horizontal Merger Guidelines* (1997) acknowledges only in the most general way that the risk of failure and loss of sunk cost investments can affect required rates of return and thus the likelihood of entry.

²An “antitrust barrier to entry” can have different meanings, as discussed in Carlton (2004), McAfee et al (2004), and Schmalensee (2004), and more generally in Viscusi et al (1995), pp. 158–164. I am following Schmalensee and using what is essentially Bain’s (1956) original definition of an antitrust barrier to entry.

distinguish among different types of risk — systematic vs. non-systematic, and aggregate (i.e., industry-wide) vs. idiosyncratic (firm-specific) — and show how they affect industry evolution differently, and thus have different antitrust implications.³

Risk can magnify sunk costs because of the opportunity cost of irreversibly investing rather than waiting for new information. When firms exercise their options to invest they give up the associated option value, which is also a sunk cost.⁴ But not all risks create option value. What matters is whether a particular type of risk leads to a symmetric or asymmetric distribution of returns; option value only arises in the latter case. As we will see, aggregate risk, whether systematic or non-systematic, leads to an asymmetric distribution of returns, while certain (but not all) idiosyncratic risks do not. For example, idiosyncratic fluctuations in firms' fixed costs lead to a symmetric distribution of returns and should not affect entry. An idiosyncratic risk of bankruptcy, however, can lead to an asymmetric distribution of returns because if a firm exits at a time when aggregate market conditions are poor, it will not quickly be replaced by a new entrant.

Caballero and Pindyck (1996) demonstrated these different effects of aggregate versus idiosyncratic risk in the context of a model of atomistic competition in which each entrant adds an incremental unit of capacity that is always utilized. Using the model, they also estimated aggregate risk and its impact on investment for two- and four-digit manufacturing industries. Novy-Marx (2007a) also derives price and output dynamics in a model of competitive entry by firms facing aggregate demand uncertainty that choose their capacities optimally, and differ in their costs of investing. He shows how this cost heterogeneity affects firms' investment decisions and the equilibrium price behavior. However, as in Caballero and Pindyck, in his model all capacity is always fully utilized.

³Posner (1979) ignores scale effects but argues that uncertainty over future cash flows can be an entry barrier by forcing entrants to bear a risk premium that makes their cost of capital higher than that of incumbents. However, he never explains what kind of uncertainty matters and why it should affect the cost of capital for entrants but not incumbents. Also, financial market structure, e.g., the extent of competition in the banking industry, can affect the cost of capital for potential entrants without affecting risk. See, e.g., Cetorelli and Strahan (2006), and Almedia and Philippon (2007). For a general treatment of the effects of sunk costs on market structure, see Sutton (1991).

⁴For a heuristic discussion of this point and its implications for antitrust analysis, see Pindyck (2008).

Grenadier (2002) and Novy-Marx (2007b) derive the Nash equilibrium investment strategies for a fixed set of n firms that can invest incrementally and face aggregate demand uncertainty. Grenadier assumes the firms are identical so that the equilibrium is symmetric, and shows how in this context industry capacity varies over time and how investment timing depends on the number of firms. Novy-Marx develops a more general model in which firms compete by adding capacity. Both authors study how firms can grow over time, but there is no entry and firms always produce at capacity.⁵

I develop a dynamic model in which firms can freely enter the industry by paying a sunk cost, which is the same for every firm. Once a firm has entered, it can produce all it wants of a homogeneous product at a constant marginal cost. I assume competition among firms currently in the industry is Nash-Cournot. Each firm knows the market demand curve when it enters the industry, but that demand curve shifts unpredictably, so there is aggregate (market-wide) risk in that future market demand is unknown. Also, that aggregate risk may be partly systematic. There are also two kinds of idiosyncratic risk. First, each firm must pay an ongoing fixed cost, which can differ across firms and will fluctuate unpredictably. Random fluctuations in fixed costs are uncorrelated across firms and are thus idiosyncratic. Second, each firm faces a risk of failure (i.e., liquidation), which can occur at any time, but is independent of what happens to other firms.

I show that sunk cost barriers to entry depend to a considerable extent on the nature and extent of the risks facing potential entrants. Each source of risk will raise barriers to entry if it affects firms' future profits asymmetrically; aggregate (market-wide) risk and the risk of failure do, but fluctuations in fixed costs do not. Most importantly, the extent and sources of risk are a basic structural feature of a market that play an important role in determining the number of firms that can be expected to enter, the timing of entry, and the evolution of market price and profit margins. The antitrust implications of these results are immediate: Evaluating market power in the context of mergers or possible anticompetitive

⁵Aguerrevere (2003) also derives the Nash equilibrium investment strategies for a fixed set of firms that invest incrementally and face aggregate demand uncertainty. He introduces time to build, and allows firms to produce at below capacity. Murti (2004) examines exit strategies for a duopoly facing a stochastically declining market demand. The model I develop here is simpler because I allow for free entry.

behavior such as collusion or predation must account for potential entry, which in turn is strongly dependent on the interaction of risk and sunk costs.⁶

Before turning to the fully dynamic model, in the next section I use a two-period example to illustrate in a simple way how risk can magnify sunk costs. Section 3 lays out and solves the full model, and shows how risk and sunk costs affect the industry growth rate and the evolution of price and margins. Section 4 develops an expression for the “full,” i.e., risk-adjusted, sunk cost of entry. Section 5 discusses the dependence of the results on the model’s parameters, and shows that for reasonable parameter values, the full sunk cost is far larger than the direct measure of sunk cost typically used in antitrust settings. Section 6 concludes.

2 A Simple Two-Period Example.

The relationship between risk and sunk cost can be illustrated with a two-period example.⁷ Consider a market with demand curve $P_t = \theta_t - Q_t$, where $\theta_1 = 10$, θ_2 will equal either $10 + \epsilon$ or $10 - \epsilon$, each with probability $\frac{1}{2}$, and $\theta_t = \theta_2$ for $t > 2$. Thus the variance of θ_2 is ϵ^2 . The market is currently served by a monopolist, but any number of additional firms can enter, at $t = 1$ or $t = 2$, by paying a sunk cost S . Let n be the number of firms that enter (so that post-entry, there are $n + 1$ firms). Because of uncertainty over θ_2 , the full sunk cost of entry will exceed the direct cost S . We want to see how n and the resulting market price P depend on S and ϵ .

Assume that entry takes no time, that any entrant can remain in the market forever with no further expenditure (i.e., there is no depreciation), that all firms have the same discount rate r , and that marginal cost is zero. Finally, I will assume that post-entry competition among firms is Cournot. Thus if n firms have entered the market, so that including the original monopolist, there are a total of $n + 1$ firms competing, each firm produces a quantity

⁶Farrell and Shapiro (1990) also examined the impact of mergers under the assumption of Cournot competition, but their’s was a static analysis that took the set of firms in the industry as fixed. Although my model is fully dynamic, I do not consider strategic aspects of sunk costs of entry that might reduce entry barriers; as Cabral and Ross (2008) have shown, sunk costs can serve as a commitment device on the part of entrants, thereby reducing the incentives for incumbent firms to try to drive entrants out by undercutting.

⁷This example is an elaboration of a model in Pindyck (2008).

$Q_i = \theta_t/(n+2)$ in each period, so that the total quantity produced and market price are $Q = \theta_t(n+1)/(n+2)$ and $P = \theta_t/(n+2)$. Also, each firm earns a profit of $\pi_i = \theta_t^2/(n+2)^2$.

To see how risk, measured by $\text{Var}(\theta_2) = \epsilon^2$, contributes to the “full” sunk cost of entry, first assume that entry can only occur only at $t = 1$, and determine the number of firms that will enter, n_1 . Taking n_1 as fixed, we will see that given a choice, these firms would prefer to wait until $t = 2$, and would then enter only if $\theta_2 = 10 + \epsilon$. The option value associated with entry is what each firm gives up (in terms of a reduction in the NPV of entry) by entering at $t = 1$ instead of waiting. To determine the contribution to sunk cost, we find the sunk cost S_2 that makes the NPV when the firms wait equal to the NPV when the firms (facing the original sunk cost S) enter at $t = 1$.

Entry in Period 1. If entry can occur only at $t = 1$, how many firms will enter? The NPV for each entrant is:

$$\begin{aligned} \text{NPV}_i^1 &= \pi_{i1} + \mathcal{E} \sum_{t=2}^{\infty} \frac{\pi_{it}}{(1+r)^{t-i}} - S \\ &= \frac{100(1+r) + \epsilon^2}{r(n+2)^2} - S \end{aligned} \quad (1)$$

Any firm can enter, so entry occurs to the point that this NPV = 0. Solving for n :

$$n_1(S, \epsilon) = \sqrt{\frac{100(1+r) + \epsilon^2}{rS}} - 2 \quad (2)$$

As expected, n_1 is decreasing in S , but because π_{i2} is a convex function of θ_2 , it is increasing in ϵ .

The Value of Waiting. Now take $n_1(S, \epsilon)$ as fixed. Because $\text{NPV}_i^1 = 0$, the firms would clearly prefer to wait until period 2 before deciding whether to enter, and would enter only if $\theta_2 = 10 + \epsilon$. The probability that $\theta_2 = 10 + \epsilon$ is .5, so if they waited, the NPV of entry for each firm (discounting back to $t = 1$) would be:

$$\text{NPV}_i^2 = \frac{(10 + \epsilon)^2}{2r(n_1 + 2)^2} - \frac{1}{2}S \quad (3)$$

Now substitute eqn. (2) for n_1 :

$$\text{NPV}_i^2 = \frac{1}{2}S \left[\frac{(10 + \epsilon)^2}{100(1+r) + \epsilon^2} - 1 \right] \quad (4)$$

This NPV is greater than zero as long as $\epsilon > 5r$ (otherwise the loss from discounting at rate r over one period exceeds the expected gain from waiting to learn about θ_2). Thus, assuming that $\epsilon > 5r$, these firms would prefer to wait until period 2 before making their entry decisions.

The Full Sunk Cost of Entry. To determine the option value associated with the sunk cost of entry, we ask what the firms give up by entering at $t = 1$ rather than waiting until $t = 2$. Equivalently, we find the sunk cost that makes the NPV of waiting until $t = 2$ equal to the zero NPV when n_1 firms, each facing a sunk cost S , enter at $t = 1$. That is, we find the cost S_2 that makes $\text{NPV}_i^2(n_1, S_2) = \text{NPV}_i^1(n_1, S) = 0$. Using eqn. (3), substituting eqn. (2) for n_1 , and rearranging gives the ratio S_2/S :

$$S_2/S = \frac{(10 + \epsilon)^2}{100(1 + r) + \epsilon^2} \quad (5)$$

Note that $S_2/S > 1$ as long as $\epsilon > 5r$.

Eqn. (5) translates the option value that is lost upon entry into an equivalent “markup” over the direct sunk cost of entry. S_2 is the full sunk cost of entry, i.e., the direct sunk cost S plus the option value that is lost by irreversibly investing. As Figure 1 illustrates (for a discount rate $r = .10$), the greater the variance of θ_2 , the greater is this lost option value, so S_2/S is increasing in ϵ . Thus the full sunk cost that is relevant to an entry decision (and therefore relevant to an analysis of entry barriers) is greater than the direct sunk cost that is typically measured, and depends on the the extent of uncertainty over future market conditions.

This example does not distinguish among different types of risk, and assumes that there is only a single change in market demand so that any entry occurs at one point in time. To explain the relationship between risk and sunk costs in more detail, we need a fully dynamic model of entry and exit.

3 A Dynamic Model of Entry and Industry Evolution.

In this section I lay out a model in which firms can enter a market by paying a sunk cost (which is the same for every firm), but entrants face both aggregate and idiosyncratic risk

on an ongoing basis. I assume that once in the market, each firm can produce any quantity it wants of a homogenous good at a constant marginal cost c , and that competition at each point in time is Cournot.⁸ The number of firms in the industry and the equilibrium market price will be determined in part by the size of the sunk cost, but also by various forms of risk.

The market demand curve is given by:

$$P_t = \theta_t - bQ_t , \tag{6}$$

where Q_t is aggregate output, and θ_t follows a geometric Brownian motion:

$$d\theta = \alpha_1\theta dt + \sigma_1\theta dz_1 . \tag{7}$$

Thus θ_t captures market or aggregate risk, which may be partly systematic, and shifts in the demand curve are parallel. (This linear demand can be viewed as an approximation to a more general demand curve.) An advantage of this demand specification is that it leads to an equilibrium threshold price that triggers new entry. Although the price fluctuates over time, in the long run it will approach an equilibrium maximum price as the number of firms grows. This makes it possible to examine how changes in various kinds of risk affect price over the long run. The same price behavior would result from a more general demand curve of the form $P_t = \theta_t - g(Q_t)$, but the resulting complexity yields no additional insights. Likewise, an alternative process could be specified for θ_t , e.g., a mean-reverting or a jump process, but the qualitative results would still hold.

To enter the industry, a firm must pay a one-time sunk cost S , and once it has entered, it can produce as much as it wants forever. (Introducing depreciation does not change the overall results.) Thus one can think of the sunk cost as providing access to a technology, or alternatively, the cost of a production facility with capacity larger than the firm's profit-

⁸Because each firm's capacity is unlimited, Bertrand competition would not yield the Cournot outcome as in Kreps and Scheinkman (1983). One could introduce differentiated products and Bertrand competition, but this would add to the complexity of the model without yielding additional insights. I also ignore the possibility of sustained collusion or cartelization, which would be unlikely given the possibility of entry. See Levenstein and Suslow (2006) for a discussion of this point.

maximizing output. All firms have the same constant marginal cost c , and an ongoing firm-specific and time-varying fixed cost $\phi_{i,t}$. Firms' fixed costs evolve as independent geometric Brownian motions; for Firm i :

$$d\phi_i = \alpha_2\phi_i dt + \sigma_2\phi_i dz_{2i} , \quad (8)$$

with $\mathcal{E}(dz_1 dz_{2i}) = 0$ for all i and $\mathcal{E}(dz_{2i} dz_{2j}) = 0$ for $i \neq j$. Note that firms' fixed costs all have the same drift and volatility, but can start from different positions, and evolve as independent stochastic processes. Thus the fixed costs capture idiosyncratic risk.⁹

There is an additional source of idiosyncratic risk: Each firm faces a risk of failure (e.g., bankruptcy leading to liquidation). I model this as a Poisson event with mean arrival rate λ ; if the event occurs, the firm's value V_i drops to zero, where it remains permanently.

3.1 Price, Markup, and Profit.

Suppose n firms are currently in the industry. Then in the Cournot equilibrium, each firm produces a quantity $Q_i = (\theta - c)/b(n + 1)$, so total output and market price are:

$$Q = \left(\frac{n}{n+1}\right) \frac{\theta - c}{b} \quad \text{and} \quad P = \frac{\theta + nc}{n+1} . \quad (9)$$

Note that the market price P is independent of the slope of the demand curve, b , and unless n changes, it moves linearly with θ_t . Also, the profit flow for each firm is:

$$\pi_{it} = \frac{(\theta_t - c)^2}{b(n+1)^2} - \phi_{it} . \quad (10)$$

From eqn. (9) for P , $(\theta - c) = (n + 1)(P - c) = (n + 1)\omega$, where $\omega \equiv P - c$ is the operating margin. Substituting this into eqn. (10), we can rewrite the firm's profit as a function of the industry-wide margin ω_t :

$$\pi_{it} = \frac{\omega_t^2}{b} - \phi_{it} . \quad (11)$$

With n firms, $\omega_t = (\theta_t - c)/(n + 1)$, so if another firm enters, the operating margin will drop to $(\theta_t - c)/(n + 2)$, i.e., ω_t will drop by an amount $\Delta\omega_n = (\theta_t - c)/(n + 1)(n + 2)$. Any potential entrant must take this into account.

⁹Idiosyncratic risk could just as easily have been introduced as shocks to productivity, as in Caballero and Pindyck (1996). What matters is that otherwise identical firms have unpredictable differences in profits.

It will be convenient to approximate the expression for the market price as follows:

$$P = \frac{\theta}{n+1} + \frac{nc}{n+1} \approx \frac{\theta}{n+1} + c, \quad (12)$$

so that $\theta \approx (n+1)(P-c) = (n+1)\omega$. This approximation is close if n is large and/or c is small. It simplifies matters because then as long as n does not change, the operating margin ω will follow the same stochastic process as θ , i.e.,

$$\frac{d\omega}{\omega} = \frac{d\theta}{\theta} = \alpha_1 dt + \sigma_1 dz_1. \quad (13)$$

When θ becomes sufficiently large, n will increase and both P and ω will drop; ω will again follow eqn. (13), but now from a (lower) starting point.

One would expect that any systematic risk is market-wide, i.e., pertains to θ_t , and that all idiosyncratic risks (i.e., random changes in ϕ_{it} or the risk of collapse) are diversifiable. Let r denote the risk-free interest rate, and $\mu \geq r$ denote the risk-adjusted expected return on θ_t (or on a portfolio of assets perfectly correlated with θ_t). An asset's return shortfall (which may or may not equal the asset's payout rate) is the asset's expected return less its expected rate of capital gain. Thus the return shortfall for θ_t is $\delta_1 = \mu - \alpha_1$, and for ϕ_{it} is $\delta_2 = r - \alpha_2$. Given eqn. (10) for each firm's profit and the processes (13) and (8) for ω_t and ϕ_{it} , we can determine the dynamics of entry, the operating margin ω_t , and the price P_t .

3.2 Entry and the Value of a Firm.

If there were n firms in the market and no further entry could occur, the value of each firm could easily be found by evaluating $\mathcal{E}_0 \int_0^\infty \pi_{it}(\theta_t, \phi_{it}) dt$, using eqn. (10) for π_{it} . With free entry, n will change, so this integral is not easily evaluated. Also, unlike in models of atomistic competition (e.g., Caballero and Pindyck (1996) or Chapter 8 of Dixit and Pindyck (1994)), the entry of a firm affects the market price, which makes the entry condition dependent on the number of firms already in the market. To determine the market dynamics, I surmise that with n firms already in the market, there is some threshold for the operating margin, $\bar{\omega}_n$, (and a price $\bar{P}_n = \bar{\omega}_n + c$) such that the $n+1$ st firm will enter whenever $\omega_t \geq \bar{\omega}_n$.

To find $\bar{\omega}_n$, I first assume that n is sufficiently large that the entry of one more firm has a negligible impact on the market price. In that case, the threshold for ω , which I denote

by $\bar{\omega}$, will not depend on n . For some arbitrary $\bar{\omega}$, I find the value of the firm when $\omega_t < \bar{\omega}$, and then find $\bar{\omega}$ and show that it is indeed an equilibrium for large n . Using the solution for $\bar{\omega}$, I return to the case of small n and determine $\bar{\omega}_n$.

3.2.1 Large n .

Suppose n is sufficiently large that entry has no significant impact on price. Consider an incumbent firm selling at the market price P_t , earning a margin $\omega_t = P_t - c$, and paying a fixed cost ϕ_i . To find the threshold $\bar{\omega}$ that triggers entry, we conjecture that $\bar{\omega}$ exists and then show that for large n , entry when ω_t reaches $\bar{\omega}$ is indeed an equilibrium.

First, suppose that some value $\bar{\omega}$ is indeed the entry threshold. If $\omega_t < \bar{\omega}$, then ω_t will surely remain below $\bar{\omega}$ during the next short interval of time. It is easily shown that the value of the firm, $V(\omega, \phi_i)$, must then satisfy:

$$\frac{1}{2}\sigma_1^2\omega^2V_{\omega\omega} + (r - \delta_1)\omega V_{\omega} + (r - \delta_2)\phi_i V_{\phi} - (r + \lambda)V + \frac{\omega^2}{b} - \phi_i = 0, \quad (14)$$

where V_{ω} denotes $\partial V/\partial\omega$, etc. This equation has the solution:

$$V(\omega, \phi_i) = B\omega^{\beta} + \frac{\omega^2}{b(2\delta_1 + \lambda - r - \sigma_1^2)} - \frac{\phi_i}{(\delta_2 + \lambda)}, \quad (15)$$

where B is yet to be determined, and β is given by:

$$\beta = \frac{1}{2} - (r - \delta_1)/\sigma_1^2 + \sqrt{\left[(r - \delta_1)/\sigma_1^2 - \frac{1}{2}\right]^2 + 2(r + \lambda)/\sigma_1^2}. \quad (16)$$

If no further entry were possible, the value of an operating firm would be given by the last two terms in eqn. (15). The first term on the right-hand side, $B\omega^{\beta}$, is the adjustment in value that is due to the potential entry of other firms.

We need to find B and the threshold $\bar{\omega}$. To find B , we use the fact that if $\bar{\omega}$ is a reflecting barrier, then $V_{\omega}(\bar{\omega}) = 0$.¹⁰ From eqn. (15),

$$V_{\omega}(\bar{\omega}) = \beta B\bar{\omega}^{\beta-1} + \frac{2\bar{\omega}}{b(2\delta_1 + \lambda - r - \sigma_1^2)} = 0, \quad (17)$$

¹⁰When ω_t reaches $\bar{\omega}$, it will surely fall, so if $V_{\omega}(\bar{\omega}) \neq 0$, an arbitrage opportunity would arise. This is a different form of the ‘‘smooth pasting’’ condition, in that it does not arise from optimality. For a discussion and derivation, see Dixit (1993), Section 3.5.

so that

$$B = -\frac{2\bar{\omega}^{2-\beta}}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)}, \quad (18)$$

and the value of an operating firm is given by:

$$V(\omega, \phi_i) = \frac{-2\bar{\omega}^{2-\beta}\omega^\beta + \beta\omega^2}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)} - \frac{\phi_i}{(\delta_2 + \lambda)}. \quad (19)$$

This value function depends on $\bar{\omega}$. To find $\bar{\omega}$ and see that it is an equilibrium threshold for large n , note that

$$V(\bar{\omega}, \phi_i) = \frac{(\beta - 2)\bar{\omega}^2}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)} - \frac{\phi_i}{(\delta_2 + \lambda)}. \quad (20)$$

If $\omega_t = \bar{\omega}$, a firm with some fixed cost ϕ_i will enter if its value, $V(\bar{\omega}, \phi_i)$, is greater than or equal to the entry cost S . But which firm will enter and with what fixed cost? Because there are an unlimited number of potential entrants, there will surely be some entrant with a fixed cost ϕ_i arbitrarily close — but not equal — to zero.¹¹ (Of course after the firm enters, its fixed cost will fluctuate.) Thus free entry ensures that:

$$V(\bar{\omega}, 0) = \frac{(\beta - 2)\bar{\omega}^2}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)} = S, \quad (21)$$

so the threshold $\bar{\omega}$ is given by:

$$\bar{\omega} = \left[\frac{\beta}{\beta - 2} b(2\delta_1 + \lambda - r - \sigma_1^2) S \right]^{\frac{1}{2}}. \quad (22)$$

To get some intuition for eqn. (22), suppose there is no uncertainty or expected growth in θ_t , so that $\sigma_1 = \lambda = \alpha_1 = 0$. In that case $\delta_1 = \mu = r$, $\beta \rightarrow \infty$, and $\bar{\omega} = (rbS)^{1/2}$. A firm that has just entered the industry has ϕ_i close to 0, so from eqn. (11), the capitalized value of its profit stream is ω^2/rb , which equals the cost of entry, S . Ignoring the fact that there must be an integer number of firms, we then have $\bar{\omega} = (rbS)^{1/2}$.

The usual risk multiplier is $\beta/(\beta - 1)$; in this case it is $\beta/(\beta - 2)$ because $\pi_{it} \sim \theta_t^2$. But note that there is no requirement that β exceed 2. If σ_1 is sufficiently large and/or δ_1 is sufficiently small, β will be less than 2 (but always greater than 1). However, $1 < \beta < 2$

¹¹We could instead assume that all firms have the same initial fixed cost, ϕ_0 , at the time of entry, but this is just equivalent to increasing S by $\phi_0/(\delta_2 + \lambda)$.

if and only if $(2\delta_1 + \lambda - r - \sigma_1^2) < 0$, so the bracketed terms in eqn. (22) are positive. Furthermore, $\bar{\omega}$ changes smoothly as $\beta \rightarrow 2$ from either direction, i.e., if δ_1 , λ , r , or σ_1 change so that $(2\delta_1 + \lambda - r - \sigma_1^2) \rightarrow 0$.¹² Likewise, $V(\omega)$ changes smoothly with β , and if $\phi_i = 0$, it is easily seen from eqn. (19) that V is always positive.

3.2.2 Small n .

If n is small, the entry of another firm will reduce the operating margin ω_t , and any potential entrant will take this into account. I posit and confirm that there is a threshold $\bar{\omega}_n = g(n)\bar{\omega}$, with $g'(n) < 0$, $g''(n) > 0$, and $g(\infty) = 1$. Recall that with n firms in the market, the Cournot equilibrium is $\omega_t = (\theta_t - c)/(n + 1)$. The $(n+1)$ st firm enters when ω_t reaches $\bar{\omega}_n$, i.e., when θ_t reaches $(n + 1)\bar{\omega}_n + c$. After entry, ω_t drops from $\bar{\omega}_n$ to $\bar{\omega}_n(n + 1)/(n + 2)$.

I show in the Appendix that $g(n)$ is given by:

$$g(n) = \left[\frac{\beta - 2}{\beta \left(\frac{n+1}{n+2}\right)^2 - 2 \left(\frac{n+1}{n+2}\right) \beta} \right]^{\frac{1}{2}}. \quad (23)$$

The Appendix also provides an expression — eqn. (34) — for $V(\omega, n)$, the value of each of the n operating firms prior to the entry of the $(n + 1)$ st firm.

Figure 2 shows $V(\omega, n)$ for $n = 1$, $n = 3$, and $n = \infty$. (The other parameters are $r = .04$, $\mu = .08$, $\alpha_1 = \lambda = 0$, $\sigma_1 = 0.2$, $b = 1$ and $S = 10$. With these parameters, $\beta = 3.56$ and $\bar{\omega} = 1.35$.) Note that if $n = 1$, a second firm will enter when ω_t reaches $\bar{\omega}_1 = 1.60$, and ω_t will then drop to $(2/3)\bar{\omega}_1 = 1.07$, at which point the value of an operating firm is equal to the sunk cost of entry, $S = 10$. For $(2/3)\bar{\omega}_1 < \omega_t \leq \bar{\omega}_2$, $V(\omega) > S$ because the incumbent can expect to enjoy a high margin for some finite time before another firm enters. As n increases, the entry threshold drops closer to $\bar{\omega}$ because the entry of another firm causes a smaller reduction in ω_t . For example, when $n = 3$, a fourth firm will enter when ω_t reaches $\bar{\omega}_3 = 1.44$, and ω_t will then drop to $(4/5)\bar{\omega}_3 = 1.15$. When n is very large, entry occurs when ω_t reaches $\bar{\omega} = 1.35$, and the resulting drop in ω_t is negligible.

Finally, note that for any given ω_t , $V(\omega, n)$ is higher when n is smaller. This is not because in the Cournot equilibrium fewer firms implies a higher operating margin; we are

¹²As $r \rightarrow 2\delta_1 + \lambda - \sigma_1^2$, $\beta \rightarrow 2$ and $\bar{\omega} \rightarrow [2(\delta_1 + \lambda + \frac{1}{2}\sigma_1^2)bS]^{1/2}$.

holding the margin fixed. The reason is that the smaller is n , the higher is the entry threshold $\bar{\omega}_n$, so the longer is the expected time before another firm enters.

3.2.3 Evolution of θ , n , and ω .

The dynamics of ω_t and n are driven by both deterministic and random changes in the demand driving variable θ_t , and by random exit at the mean rate λ . With n firms in the market and $\theta_t < (n+1)\bar{\omega}_n + c$, no further entry occurs until θ_t reaches $(n+1)\bar{\omega}_n + c$ (so that ω_t reaches $\bar{\omega}_n$), or until an incumbent firm exits, making room for an entrant. As n gets larger, each drop in ω_t becomes smaller, and ω_t stays closer to $\bar{\omega}$ on average. Also, as n increases the average time until ω_t reaches $\bar{\omega}_n$ and another firm enters falls.¹³

Figures 3 and 4 show sample paths for θ_t , the number of firms n , and the operating margin ω_t . For both sample paths, $r = .04$, $\mu = .08$, $\alpha_1 = 0$, $\sigma_1 = 0.2$, $c = 1$, and $S = 10$. In Figure 3, $\lambda = 0$, so that there is no exit, and $\theta_0 = 3$ so that initially there is one firm in the market. In Figure 4, $\lambda = 0.1$, so that each active firm has a 10-percent chance of exiting each year. With $\lambda > 0$, n will grow if θ_t increases sufficiently, but will also drop at random times as incumbent firms exit. Also, if θ_t drops sufficiently so that ω_t is well below $\bar{\omega}_n$, when one or more firms “collapse” and exit, other firms will not enter until θ_t increases sufficiently to bring ω_t back to the entry threshold.

Note from Figure 3 that the entry threshold $\bar{\omega}_n$ falls as the number of firms increases. The entry of the second firm occurs after about 1-1/2 years, when θ_t increases sufficiently so that ω_t reaches $\bar{\omega}_1 \approx 1.4$. Then ω_t drops to $(2/3)\bar{\omega}_1 = 0.93$, and the new threshold becomes $\bar{\omega}_2 \approx 1.3$. After about 15 years there are 8 firms in the market, but θ_t and ω_t drop substantially so that no further entry occurs. In Figure 4, the initial value of θ_t is set at 5.0 (corresponding to $\omega = 2$), so that we again begin with one firm. After 2 years θ_t and ω_t are high enough so that a second firm enters. (Note that $\bar{\omega}_1$ is now 2.8, about twice its value when λ was zero. A higher λ increases both $\bar{\omega}$ and β , which increases $g(n)$.) After another

¹³After the entry of the n th firm, ω_t drops from $\bar{\omega}_n = g(n)\bar{\omega}$ to $\bar{\omega}_n(n+1)/(n+2)$, and the $(n+1)$ st firm enters when ω_t increases to $\bar{\omega}_{n+1} = g(n+1)\bar{\omega}$. Thus as n becomes larger, the percentage increase in ω_t needed to induce entry of another firm, $\Delta_n \equiv \Delta \log \omega_t = \log[g(n+1)/g(n)] - \log[(n+1)/(n+2)]$, falls, as does the expected time for this increase to occur.

year, a firm exits, but θ_t and ω_t are still sufficiently high that it is immediately replaced by another entrant, and shortly thereafter, a third firm enters. Around year 6, a firm exits, but ω_t has declined so that it is not immediately replaced. The number of firms reaches 5, but then declines to 2 as exit occurs but ω_t remains too low to induce further entry.

3.3 Industry Growth Rate.

The sample paths in Figures 3 and 4 are illustrative, but they do not address the question of how sunk costs and risk affect the expected rate of industry growth. For example, suppose that initially demand is such that there is one firm in the industry. Then, for a given set of parameter values, what is the expected value of the time it will take until there are $n > 1$ active firms? Alternatively, what is the expected value of the number of firms that will be in the industry after T years? I address these questions below.

3.3.1 Expected Time to Reach n Firms.

Suppose that at $t = 0$, $\theta_0 = m\bar{\omega}_{m-1} + c$ so that $\omega_0 = \bar{\omega}_{m-1}$, and it is just profitable for the m th firm to be in the market. Let ${}_mT_n$ be the time until there are $n > m$ firms in the market, i.e., the time until θ_t increases to $n\bar{\omega}_{n-1} + c$ so that ω_t reaches $\bar{\omega}_{n-1}$, triggering the entry of the n th firm. Let $\mathcal{E}({}_mT_n)$ be the expectation of ${}_mT_n$. Define $\alpha'_1 \equiv \alpha_1 - \frac{1}{2}\sigma_1^2$. Using Theorem 5.3 in Karlin and Taylor (1975), as long as $\alpha'_1 > 0$, $\mathcal{E}({}_mT_n)$ is given by:

$$\mathcal{E}({}_mT_n) = \frac{1}{\alpha'_1} \log \left[\frac{n\bar{\omega}_{n-1} + c}{m\bar{\omega}_{m-1} + c} \right]. \quad (24)$$

Note that $\mathcal{E}({}_nT_{n+1}) \rightarrow 0$ as $n \rightarrow \infty$, i.e., as the industry becomes larger, on average new firms enter more quickly.

An increase in σ_1 has two opposing effects on $\mathcal{E}({}_mT_n)$. First, it increases each $\bar{\omega}_k$, $k = m, m + 1, \dots, n - 1$, so that θ_t must rise by a larger amount to trigger entry. Second, it increases the volatility of θ_t , increasing the probability that θ_t will reach any given limit in a fixed time interval. With some algebra, it can be shown that the first effect dominates, so that $(d/d\sigma_1)\mathcal{E}({}_mT_n) > 0$. For example, suppose the parameters are the same as for the sample path in Figure 3 (so that $\sigma_1 = 0.2$), except set $\alpha_1 = .20$ to ensure that $\alpha'_1 > 0$. If we

start with 1 firm, the expected time to reach 5 firms is $\mathcal{E}({}_1T_5) = 6.29$ years. If σ_1 is increased to 0.4, $\mathcal{E}({}_1T_5)$ increases to 23.19 years. This is just another manifestation of industry-wide risk magnifying the direct sunk cost of entry, so that greater risk requires a larger expansion of market demand to induce entry of any fixed number of firms.

If $\alpha'_1 < 0$, ${}_mT_n$ has a defective probability distribution, i.e., there is a positive probability that ${}_mT_n$ will be infinite. However, in this case one can still determine (via Monte Carlo simulation) the expected time until there are n firms *given* that this time is finite. This is done by generating sample paths for θ_t over some arbitrarily long horizon (e.g., 200 years), and then averaging over only those paths for which θ_t reached $\bar{\omega}_{n-1} + c$ within the horizon.

3.3.2 Expected Number of Firms.

Assume once again that at $t = 0$, $\theta_0 = m\bar{\omega}_{m-1} + c$ so that $\omega_0 = \bar{\omega}_{m-1}$, and thus it is just profitable for the m th firm to be in the market. Let n be the number of firms at $t = T$. If $\lambda = 0$, then $n \geq m$ is also the *maximum* number of firms in the market between $t = 0$ and T , but if $\lambda > 0$, firms can exit so n may be less than the maximum, and may also be less than the starting number of firms, m . In either case, one can find the expected number of firms in the market at T , $\mathcal{E}(n|m, T)$, by simple Monte Carlo simulation.

If $\alpha'_1 \equiv \alpha_1 - \frac{1}{2}\sigma_1^2 < 0$, we can also find the expected value of the *maximum* number of firms that will ever be in the market. Let $x_t = \log \theta_t$ and let $\bar{x} = \max_{0 \leq t < \infty} (x_t - x_0)$. Then \bar{x} has the exponential distribution:

$$\Pr(\bar{x} \geq x) = e^{-\xi x}$$

where $\xi = 2|\alpha'_1|/\sigma_1^2$. (See Karlin and Taylor, Corollary 5.1.) Since θ_t determines the maximum number of firms (for any value of λ), we can use this to get the distribution for the maximum number of firms given that there are initially m firms, $\bar{n}_m \equiv \max_{0 \leq t < \infty} n_t | n_0 = m$:

$$\begin{aligned} \Pr(\bar{n}_m \geq n) &= \Pr[\bar{x} \geq \log(n\bar{\omega}_{n-1} + c) | x_0 = \log(m\bar{\omega}_{m-1} + c)] \\ &= \Pr\left[\bar{x} \geq \log\left(\frac{n\bar{\omega}_{n-1} + c}{m\bar{\omega}_{m-1} + c}\right)\right] \\ &= \exp\left[-\xi \log\left(\frac{n\bar{\omega}_{n-1} + c}{m\bar{\omega}_{m-1} + c}\right)\right] = \left(\frac{m\bar{\omega}_{m-1} + c}{n\bar{\omega}_{n-1} + c}\right)^\xi \end{aligned} \quad (25)$$

As one would expect, $\partial \Pr(\bar{n}_m \geq n)/\partial n < 0$ and $\Pr(\bar{n}_m \geq n) \rightarrow 0$ as $n \rightarrow \infty$. Also, it is easy to show that for any $n > m$, $\partial \Pr(\bar{n}_m \geq n)/\partial \sigma_1 > 0$ and $\partial \Pr(\bar{n}_m \geq n)/\partial |\alpha'_1| < 0$. Using (25), one can compute the expected value of the maximum number of firms given that there are m firms currently in the market, $\mathcal{E}(\bar{n}_m)$.

Table 1 shows, for $\sigma_1 = .2$ and $.4$ and $\lambda = 0$ and $.1$, the expected number of firms after 5 years and after 10 years, when there is initially one firm in the market. For either time horizon, the expected number of firms is increasing in σ_1 and decreasing in λ . $\mathcal{E}(n|1, T)$ is increasing in σ_1 because $\alpha'_1 < 0$ (i.e., the percentage drift of θ_t is negative), so a higher value of σ_1 increases the probability of $\theta_T - \theta_0 > \Delta$ for any positive Δ . The table also shows the expected value of the maximum number of firms that will ever be in the market. Note that if $\lambda = .1$, $\mathcal{E}(n|1, 10) < \mathcal{E}(n|1, 5)$ because the average rate of exit exceeds the average rate of demand growth-induced entry.

3.4 Determinants of the Entry Threshold.

I turn now to the entry threshold $\bar{\omega}_n$ and examine its dependence on the various parameters of the model. Recall that $\bar{\omega}_n = g(n)\bar{\omega}$, where $\bar{\omega}$ is the entry threshold when the number of firms n is very large, and $g(n) > 1$ is the increase in the threshold resulting from the post-entry drop in ω_t when n is small. It is easiest to discuss $\bar{\omega}$ and $g(n)$, both of which depend on β , separately. I begin with $g(n, \beta) = \bar{\omega}_n/\bar{\omega}$, which is given by eqn. (23). First, note that $\partial g(n, \beta)/\partial \beta > 0$, and recall from eqn. (16) that $\beta > 1$ and β is decreasing in σ_1 and increasing in δ_1 . Other things equal, if n is small, a lower value of σ_1 and/or higher value of $\delta_1 = \mu - \alpha_1$ implies a larger difference between $\bar{\omega}_n$ and $\bar{\omega}$. The reason is that the drop in ω_t to $\bar{\omega}_n(n+1)/(n+2)$ that will occur following entry will be relatively permanent, because the volatility σ_1 and drift α_1 of θ_t are low. Thus the entrant requires a post-entry value of ω_t that is higher than would be the case otherwise, which implies a higher threshold.

Second, as with eqn. (22), there is no requirement that β be greater than 2. If $1 < \beta < 2$, both the numerator and denominator in eqn. (23) will be negative. Also, as β approaches 2 from either direction, $g(n, \beta) \rightarrow (\frac{n+2}{n+1})[1 + 2 \ln(\frac{n+2}{n+1})]$.

Now consider $\bar{\omega}$, the threshold for large n . Note that $\bar{\omega}$ does not depend on σ_2 , the

To summarize, $\partial\bar{w}/\partial\sigma_1 > 0$, $\partial\bar{w}/\partial\lambda > 0$, $\partial\bar{w}/\partial\mu > 0$, and $\partial\bar{w}/\partial r < (>)0$ if $r \leq (>)\mu$. Of course we are assuming that changes in σ_1 , μ , etc. are made holding all other parameters fixed, and as discussed later, that is not always the right thought experiment.

4 The Magnification of Sunk Costs.

One way to measure the effect of risk on market structure is by determining how it “magnifies” the sunk cost of entry, S . To begin, I focus on the “large- n ” threshold, \bar{w} . Suppose $\sigma_1 = \lambda = 0$ so that $\beta \rightarrow \infty$ and $\beta/(\beta - 2) \rightarrow 1$. In this “riskless world,” \bar{w} is smaller than it would be if σ_1 or λ took on their actual values. Thus I find the sunk cost $S^* > S$ that makes \bar{w} when $\sigma_1 = \lambda = 0$ equal to what it was for the given values of σ_1 , λ , and S . From eqn. (22) for \bar{w} , S^* therefore satisfies the equation:

$$b(2\delta_1^* - r)S^* = \left(\frac{\beta}{\beta - 2}\right)b(2\delta_1 + \lambda - r - \sigma_1^2)S, \quad (26)$$

where $\delta_1^* = \mu^* - \alpha_1$ is the value of δ_1 that corresponds to the riskless world in which $\sigma_1 = \lambda = 0$. With no risk, $\mu^* = r$ so $\delta_1^* = r - \alpha_1$.

We can then calculate the “markup” given by:

$$S^*/S = \left(\frac{\beta}{\beta - 2}\right)\frac{(2\delta_1 + \lambda - r - \sigma_1^2)}{(2\delta_1^* - r)} = \left(\frac{\beta}{\beta - 2}\right)\frac{(2\mu - 2\alpha_1 + \lambda - r - \sigma_1^2)}{(r - 2\alpha_1)}. \quad (27)$$

This markup is independent of b , the slope of the demand curve. It is easy to see that $S^* \geq S$ is increasing in σ_1 , and if $\sigma_1 = \lambda = 0$ and $\mu = r$, $S^* = S$.¹⁵ Note from eqns. (16) and (27) that although $\partial\beta/\partial\lambda > 0$, $\partial(S^*/S)/\partial\lambda > 0$. Both σ_1 and λ embody risk that asymmetrically impacts a firm’s value, and an increase in either magnifies the sunk cost of entry.¹⁶

To determine S^*/S for small n , note that if $\sigma_1 = \lambda = 0$, $\beta \rightarrow \infty$ and, from eqn. (23), $g(n) \rightarrow (n + 2)/(n + 1)$, so that $\bar{w}_n \rightarrow \left(\frac{n+2}{n+1}\right)[b(2\delta_1^* - r)S]^{1/2}$. Thus, eqn. (26) becomes:

$$\left(\frac{n+2}{n+1}\right)^2 b(2\delta_1^* - r)S^* = g^2(n) \left(\frac{\beta}{\beta - 2}\right)b(2\delta_1 + \lambda - r - \sigma_1^2)S. \quad (28)$$

¹⁵If $\sigma_1 = \lambda = 0$, there is no risk, so we would expect $\mu = r$. If $\sigma_1 = \lambda = 0$ and $\mu > r$, then $S^*/S > 1$. Also, note that S^*/S is defined only for $r > 2\alpha_1$; if $r \leq 2\alpha_1$, the value of an operating firm in the absence of risk is infinite.

¹⁶Also, note that $S^*/S > 0$, because $1 < \beta < 2$ iff $(2\delta_1 + \lambda - r - \sigma_1^2) < 0$.

Rearranging and substituting eqn. (23) for $g(n)$, the “markup” is now:

$$S^*/S = \left[\frac{\beta}{\beta - 2\left(\frac{n+1}{n+2}\right)^{\beta-2}} \right] \frac{(2\delta_1 + \lambda - r - \sigma_1^2)}{(2\delta_1^* - r)}. \quad (29)$$

With some algebra, it can be shown that S^*/S is decreasing in n . The reason is that a larger n implies a lower entry threshold $\bar{\omega}_n$ and a smaller reduction in ω_t when another firm enters, increasing the value function for each operating firm and thus increasing the net payoff from entry, $V(\omega, n+1) - S$. This is equivalent to a lower S^* in the “riskless world.”

For any number of firms n , eqn. (29) shows how market risk as measured by σ_1 and the idiosyncratic risk of collapse, λ , affect the markup S^*/S . Of interest is the different effects of systematic versus non-systematic risk. To address this, we decompose the markup into two components. By the CAPM, $\mu = r + \phi\rho_{\theta m}\sigma_1$, where ϕ is the market price of risk, and $\rho_{\theta m}$ is the correlation of θ_t with the market. Thus if $\rho_{\theta m}$ doesn’t change, μ (and therefore δ_1) should increase linearly with σ_1 . In particular, $\phi = (r_m - r)/\sigma_m$, where r_m is the expected return on the market and σ_m is the standard deviation of that return. Using $r_m - r \approx .08$ and $\sigma_m \approx 0.2$, $\phi \approx 0.4$. If, e.g., $\rho_{\theta m} = 0.5$, then $\mu = r + 0.2\sigma_1$.

By assumption the risk of collapse, λ , is non-systematic. We can therefore consider two cases: (1) $\rho_{\theta m} = 0$, so that all risk is non-systematic; and (2) $\rho_{\theta m} = 1$, so that all industry-wide risk is systematic, i.e., stochastic movements in θ_t are perfectly correlated with the overall market. Then for any given value of σ_1 , we find S^*/S under the assumption that $\rho_{\theta m} = 0$, which implies that $\mu = r$ so that $\delta_1 = \mu - \alpha_1 = r - \alpha_1$, and denote this ratio by S_n^*/S . (Note that δ_1 appears directly in the expression for S^*/S and indirectly via eqn. (16) for β .) Likewise, we find S^*/S under the assumption that $\rho_{\theta m} = 1$, which implies that $\mu = r + \phi\sigma_1$, and denote this ratio by S_s^*/S . Thus $(S_s^* - S_n^*)/S \geq 0$ measures the *maximum* differential impact of systematic risk. Note that if S^*/S is the ratio for the true value of $\rho_{\theta m}$, then $(S^* - S_n^*)/S$ is the differential contribution of systematic risk for the actual industry.

As an example, take the parameter values used to generate the sample paths in Figures 3 and 4: $r = .04$, $\mu = .08$, $\alpha_1 = 0$, $\sigma_1 = 0.2$, $c = 1$, and $S = 10$. Assuming n is large, with $\lambda = 0$, $S^*/S = 4.63$, $S_n^*/S = 3.05$, and $S_s^*/S = 6.47$. Thus the differential impact of systematic risk on the sunk cost markup is $4.63 - 3.05 = 1.58$, i.e., it increases the markup by about 52%.

Also, the maximum differential impact of systematic risk is $6.47 - 3.05 = 3.42$, which more than doubles the markup. If $\lambda = 0.1$, the markups are all much larger (because there is a risk of “collapse” in addition to market demand risk). Then $S^*/S = 8.25$, $S_n^*/S = 6.86$, and $S_s^*/S = 9.88$, so the differential impact of systematic risk is 1.39.

5 Comparative Statics.

How do changes in various parameters affect the limiting operating margin $\bar{\omega}$, the markup on sunk cost S^*/S , and the maximum differential markup $(S_s^* - S_n^*)/S$? This question is not straightforward because a change in one parameter is likely to be accompanied by a change in one or more other parameters. For example, we saw that an increase in σ_1 — holding all other parameters fixed — results in increases in both $\bar{\omega}$ and S^*/S . But should all other parameters be held fixed?

5.1 Relationships Among Parameters.

As explained earlier, changes in σ_1 should affect μ , the risk-adjusted expected rate of return on θ_t , through the CAPM: $\mu = r + \phi\rho_{\theta m}\sigma_1$, where $\phi = (r_m - r)/\sigma_m$ is the market price of risk. I will use $\rho_{\theta m} = 0.5$, so with $r_m - r \approx .08$ and $\sigma_m \approx 0.2$, $\phi \approx 0.4$ and $\mu = r + 0.2\sigma_1$.

Likewise, one might expect a change in the failure rate λ to be accompanied by a change in α_1 . Suppose the risk-adjusted expected rate of return μ is independent of λ (which would be the case if Poisson failures are non-systematic). This expected rate of return is the sum of an expected rate of capital gain, $\alpha_1 - \lambda$, and a payout rate (or return shortfall). To keep the expected rate of capital gain and thus the expected return fixed, α_1 must increase with λ , i.e., $\alpha_1 - \lambda$ should be constant, so that $\alpha_1 = \alpha_{10} + \lambda$.¹⁷ (Note that in this case, $\delta_1 = \mu - \alpha_1 + \lambda = \mu - \alpha_{10}$ would also remain constant.)

¹⁷See Section 5B of Dixit and Pindyck (1994) for a more detailed discussion of this point. One might also want α_1 to vary with σ_1 . Because θ_t follows a GBM, its expected percentage rate of growth is $\alpha_1 - \frac{1}{2}\sigma_1^2$ (the drift of $\log \theta_t$), so to keep this expected rate of growth at zero, we must set $\alpha_1 = \frac{1}{2}\sigma_1^2$. However, in that case δ_1 will be a non-monotonic function of σ_1 , first increasing, then decreasing and becoming negative (at which point there would be no investment) for large values of σ_1 . Also, $\mathcal{E}(\theta_t) = \theta_0 e^{\alpha_1 t}$. Thus in what follows, I do not constrain the growth rate of demand.

A change in the risk-adjusted expected return, μ could occur without an accompanying change in σ_1 (or the risk-free rate, r) if $\rho_{\theta m}$ were to increase. But as explained earlier, a change in μ could also result from a change in σ_1 through the CAPM. An increase in μ by itself, like an increase in λ , has two opposing effects. First, it will increase δ_1 and thereby increase β and reduce $\bar{\omega}$. Second, as can be seen from the second term on the RHS of eqn. (15), the increase in δ_1 also reduces the present value of future profits. This second effect dominates, so an increase in μ reduces $V(\omega_t)$, increasing both $\bar{\omega}$ and S^*/S .

For purposes of comparison, it is useful to see how the entry threshold and sunk cost markup change as individual parameters are changed while holding the other parameters fixed. Thus I present parameter-by-parameter results below. I also show the effects of parameter changes with the single constraint $\mu = r + \phi\rho_{\theta m}\sigma_1$, and with the additional constraint $\alpha_1 = \alpha_{10} + \lambda$.

5.2 Parameter Values.

What are “reasonable” values for the parameters of the model? Because entry depends only on the margin ω_t , the threshold $\bar{\omega}_n$ and sunk cost markup S^*/S are independent of the marginal cost c . The slope of the demand curve, b , does not affect S^*/S ; it does affect $\bar{\omega}_n$, but note from eqn. (22) that it simply “scales” the direct sunk cost S . Thus I set $b = 1$ and $S = 10$. I also set α_1 (and α_{10}), the drift of the demand driving variable θ_t , to zero.

Reasonable values for the risk-free rate r could be in the range of .02 to .06, and the risk premium $\mu - r$ could be in the range of 0 to .10. The average exit rate λ varies considerably across industries; I will consider values from 0 to 0.2.

That leaves aggregate demand volatility, σ_1 ($\bar{\omega}_n$ and S^*/S are unaffected by the idiosyncratic volatility of fixed costs, σ_2). Caballero and Pindyck (1996) estimated sample standard deviations of the log of the marginal profitability of capital (Π_K) for 2-digit and 4-digit SIC industries over the 29-year period 1958–1986, and those numbers are roughly comparable to σ_1 . Note from eqn. (9) that for an industry in my model, $\Pi_K = P - c = (\theta - c)/(n + 1)$, so during periods when n does not change, $d(\log \Pi_K) = (d\theta - c)/(\theta - c)$. Thus if c is small

relative to θ (*not* relative to P), $d(\log \Pi_K) \approx d\theta/\theta$.¹⁸ The 2-digit (4-digit) standard deviations range from .06 to .25 (.22 to .51), which suggests a range for σ_1 of roughly 0.1 to 0.5. This range is also consistent with the 20% annual standard deviation of real returns on the NYSE Index, which when unlevered (using an average debt-equity ratio of 1) becomes 10%. The standard deviations of unlevered individual company returns are substantially higher, and vary considerably across industries.¹⁹

For the comparative static results below, I assume that there are currently 5 firms in the market, and I use the following “base case” parameter values: $r = .04$, $\mu = .08$, $\alpha_1 = \alpha_{10} = 0$, $\sigma_1 = 0.2$, $\rho_{\theta m} = 0.5$ so that $\phi\rho_{\theta m} = 0.2$, and $\lambda = 0$. I vary these parameters within the ranges described above.

5.3 Results.

Table 2 shows how $\bar{\omega}_n$ and S^*/S vary with σ_1 , μ , λ , and r , first holding all other parameters fixed (Part A), then constraining $\mu = r + \phi\rho_{\theta m}\sigma_1 = r + 0.2\sigma_1$ (Part B), and finally adding the constraint $\alpha_1 = \alpha_{10} + \lambda$ (Part C). In all cases, $n = 5$. Note from Part A, that the markup S^*/S and the maximum differential impact of systematic risk, $\Delta S^*/S = (S_s^* - S_n^*)/S$, rise sharply as σ_1 is increased. ($S^*/S > 1$ when $\sigma_1 = 0$ because μ is fixed at .08, reducing the value function.) If μ is constrained by the CAPM, as in Parts B and C, $S^*/S = 1$ when $\sigma_1 = 0$, but $\bar{\omega}$ and S^*/S rise even more rapidly as σ_1 is increased, because the higher μ that accompanies a higher σ_1 implies a higher δ_1 and larger reduction in the value function.

Given that values of σ_1 around 0.2 to 0.6 are plausible, these results show that aggregate risk can have very large impact on equilibrium entry and market evolution. It also means that the *full* sunk cost of entry is much larger than the direct sunk cost that is usually measured and discussed in antitrust contexts.

The table also shows how increases in μ (holding everything else fixed) increase $\bar{\omega}$ and

¹⁸If $c = .5P$, $\theta = (.5n + 1)P$, so if, say, $n = 5$, $\theta = 3.5P = 7c$, making the approximation fairly close.

¹⁹Franco and Philippon (2007) examine sales and employment for 526 firms over the period 1970 to 2002, and find that permanent shocks (to productivity and relative demands) are mostly uncorrelated across firms, and by that measure are idiosyncratic. This is not unexpected, however, because the firms span the entire economy. Aggregate shocks in my model are largely industry-specific as opposed to economy-wide.

S^*/S . This is due largely to greater discounting of future profits. If μ is fixed (as in Part A), increases in r have the opposite effect; future profits are discounted at the same rate irrespective of r , but the (non-stochastic) direct sunk cost of future entry is discounted at a higher rate.

Finally, If α_1 is held fixed, as in Parts A and B, increases in the failure rate, λ , have a large impact on $\bar{\omega}$ and S^*/S . This impact is reduced somewhat if α_1 increases with λ so that the expected rate of capital gain on ω_t remains fixed, as in Part C. But the effect remains substantial, because it works largely through the direct reduction of the value function.

In summary, market-wide risk (as captured by σ_1) and the idiosyncratic risk of failure (as captured by λ) strongly affect the entry threshold and sharply increase the “full” sunk cost of entry. This is true even if market-wide risk is completely non-systematic. Thus, focusing on how risk affects firms’ cost of capital, as is typically done in antitrust settings, can be highly misleading and lead to a gross underestimate of the effects of risk on entry.

6 Conclusions.

The model developed here is quite simple in many respects. For example, all firms, including potential entrants, are assumed to be identical, with the same constant marginal cost c , and the same direct sunk cost of entry S . I assumed Cournot competition among incumbent firms and ignored possible strategic interactions leading, e.g., to implicit collusion. Also, the risk of failure (as measured by the Poisson arrival rate λ) is assumed constant and independent of market-wide conditions (e.g., ω_t) or firm-specific conditions (e.g., the firm’s current fixed cost ϕ_{it}). These assumptions are analytically convenient, but still provide a framework that elucidates how different kinds of risk affect industry dynamics and concentration by magnifying the sunk costs of entry. I have focused on a measure of the “full,” i.e., risk-adjusted, sunk cost of entry, and shown that for reasonable parameter values, this full sunk cost is far larger than the direct measure of sunk cost typically used in antitrust settings.

Market-wide risk (as captured by stochastic fluctuations in market demand) as well as the idiosyncratic risk of failure and exit both asymmetrically affect the post-entry value of

a firm, and thereby increase entry thresholds by magnifying the direct sunk cost of entry. I have argued that “reasonable” values for σ_1 are in the range of 0.2 to 0.6. As shown in Table 2, whether σ_1 is closer to 0.2 or 0.6 makes a considerable difference for the entry threshold $\bar{\omega}$ and the sunk cost markup S^*/S . It also matters how much of this market-wide risk is systematic versus non-systematic. But even if the market-wide risk is entirely non-systematic, it can dramatically increase S^*/S , even though it will have no effect on firms’ costs of capital.

The entry threshold $\bar{\omega}$ and sunk cost markup S^*/S are also sensitive to the failure rate, λ . Increases in λ cause substantial increases in both $\bar{\omega}$ and S^*/S . This is the case even though I have assumed that the risk of failure is completely non-systematic. (Reworking the model so $\lambda = \lambda(\omega_t)$ is a natural way to introduce systematic risk into the failure rate, but complicates the solution.)

I have treated risk as a basic structural feature of a market. Although it has not been my focus, the model can be used to study the effects of risk on market evolution, including the rate of growth of the number of firms. I have shown how the model can be used to compute the expected time for an industry to grow from m to $n > m$ firms, $\mathcal{E}(mT_n)$, and the expected number of firms at a future time T , $\mathcal{E}(n|m, T)$. We have seen that $\mathcal{E}(mT_n)$ is increasing with σ_1 ; this is just another example of how industry-wide risk magnifies the direct sunk cost of entry, requiring a larger expansion of market demand to induce entry.

These results suggest that in antitrust settings, the extent and nature of market-wide risk and the risk of failure should be of central concern. It is common in antitrust analyses to focus on how risk affects firms’ cost of capital, but this is very misleading. We have seen that risk might have no effect on the cost of capital but can still act as an entry barrier.

Appendix A: Entry Threshold for Small n

Value of an Operating Firm. When n is small, B in eqn. (15) depends on n and thus is no longer given by eqn. (18). With n firms in the market, the reflecting barrier is $\bar{\omega}_n$, so the smooth pasting condition is $V_\omega(\bar{\omega}_n) = 0$, and $B(n)$ is given by:

$$B(n) = -\frac{2\bar{\omega}_n^{2-\beta}}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)}. \quad (30)$$

Thus with n firms, the value of an operating firm is:

$$V(\omega, n) = -\frac{2\bar{\omega}_n^{2-\beta}}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)}\omega_t^\beta + \frac{\omega_t^2}{b(2\delta_1 + \lambda - r - \sigma_1^2)}. \quad (31)$$

After ω_t reaches $\bar{\omega}_n$, the number of firms increases to $n+1$, so ω_t drops to $\bar{\omega}_n(n+1)/(n+2)$, and the new entry threshold becomes $\bar{\omega}_{n+1}$. The value of each operating firm (including the one that just entered) therefore becomes $V(\frac{n+1}{n+2}\bar{\omega}_n, n+1)$. To determine $g(n)$ and $\bar{\omega}_n$, I approximate $V(\frac{n+1}{n+2}\bar{\omega}_n, n+1)$ by $V(\frac{n+1}{n+2}\bar{\omega}_n, n)$:

$$V(\frac{n+1}{n+2}\bar{\omega}_n, n) = \frac{\beta(\frac{n+1}{n+2})^2 - 2(\frac{n+1}{n+2})^\beta}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)}\bar{\omega}_n^2. \quad (32)$$

I show below that for $n \geq 2$, the approximation error is small.²⁰

Entry Threshold. We can now determine the function $g(n)$ and thus the entry threshold $\bar{\omega}_n$. Entry will occur when the post-entry value is equal to the entry cost S . Setting the post-entry value from eqn. (32) equal to S and substituting $g(n)\bar{\omega}$ for $\bar{\omega}_n$ gives

$$g(n)\bar{\omega} = \left[\frac{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)S}{\beta(\frac{n+1}{n+2})^2 - 2(\frac{n+1}{n+2})^\beta} \right]^{\frac{1}{2}}. \quad (33)$$

Now use eqn. (??) for $\bar{\omega}$ to obtain eqn. (23) for $g(n)$. Note that as expected, $g(n) > 1$, $g'(n) < 0$, $g''(n) > 0$, and $g(\infty) = 1$. Also, note that after the entry of the $(n+1)$ st firm, ω_t drops from $\bar{\omega}_n > \bar{\omega}$ to $\bar{\omega}_n(n+1)/(n+2) < \bar{\omega}$.

We also want the value of each of the n operating firms prior to the entry of the $(n+1)$ st firm. Using eqn. (31), for $\omega_t \leq g(n)\bar{\omega}$, $V(\omega, n)$ is:

$$V(\omega, n) = -\frac{2[g(n)\bar{\omega}]^{2-\beta}\omega_t^\beta}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)} + \frac{\omega_t^2}{b(2\delta_1 + \lambda - r - \sigma_1^2)}. \quad (34)$$

²⁰Without this approximation, finding $g(n)$ requires solving a nonlinear difference equation, which must be done numerically.

Approximation Error. The entry condition is $V(\frac{n+1}{n+2}\bar{\omega}_n, n+1) = S$, which I have approximated with $V(\frac{n+1}{n+2}\bar{\omega}_n, n) = S$. The approximation error is therefore $\Delta V_n = [-B(n) + B(n+1)]\bar{\omega}_n^\beta$. Using eqn. (30), this error is:

$$\Delta V_n = \frac{2}{\beta b(2\delta_1 + \lambda - r - \sigma_1^2)} [\bar{\omega}_n^2 - \bar{\omega}_{n+1}^{2-\beta} \bar{\omega}_n^\beta]$$

Using $\bar{\omega}_{n+1} = \bar{\omega}_n g(n+1)/g(n)$, the percent error can be written as:

$$\frac{\Delta V_n}{V(\bar{\omega}_n, n)} = \frac{2 - 2 \left[\frac{g(n)}{g(n+1)} \right]^{\beta-2}}{\beta \left(\frac{(n+1)}{(n+2)} \right)^2 - 2 \left(\frac{(n+1)}{(n+2)} \right)^\beta}.$$

To estimate the percent error, use eqn. (23) for $g(n)$, so that:

$$\frac{g(n)}{g(n+1)} = \frac{(n+2)^2}{(n+1)(n+3)} \left[\frac{\beta - 2 \left(\frac{n+2}{n+3} \right)^{\beta-2}}{\beta - 2 \left(\frac{n+1}{n+2} \right)^{\beta-2}} \right]^{1/2}.$$

If $\beta = 3$ ($\beta = 5$), the percent error is 3.7% (5.8%) when $n = 3$, 1.4% (2.2%) when $n = 5$, and 0.3% (0.5%) when $n = 10$.

References

- Aguerrevere, Felipe L., “Equilibrium Investment Strategies and Output Price Behavior: A Real Options Approach,” *Review of Financial Studies*, Winter 2003, **16**, 1239–72.
- Almeida, Heitor, and Thomas Philippon, “The Risk-Adjusted Cost of Financial Distress,” *Journal of Finance*, December 2007.
- Bain, Joe S., *Barriers to New Competition*, Harvard University Press, 1956.
- Caballero, Ricardo J., and Robert S. Pindyck, “Uncertainty, Investment, and Industry Evolution,” *International Economic Review*, August 1996, **37**, 641–62.
- Cabral, Luis M.B., and Thomas W. Ross, “Are Sunk Costs a Barrier to Entry?” *Journal of Economics & Management Strategy*, Spring 2008, **17**, 97–112.
- Carlton, Dennis W., “Why Barriers to Entry are Barriers to Understanding,” *American Economic Review*, May 2004, **94**.
- Cetorelli, Nicola and Philip E. Strahan, “Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local U.S. Markets,” *Journal of Finance*, February 2006, **61**, 437–61.
- Dixit, Avinash K., *The Art of Smooth Pasting*, Harwood Academic Publishers, 1993.
- Dixit, Avinash K., and Robert S. Pindyck, *Investment Under Uncertainty*, Princeton University Press, 1994.
- Farrell Joseph, and Carl Shapiro, “Horizontal Mergers: An Equilibrium Analysis,” *American Economic Review*, March 1990, **80**, 107–126.
- Franco, Francesco, and Thomas Philippon, “Firms and Aggregate Dynamics,” *Review of Economics and Statistics*, Nov. 2007, **89**, 587–600.
- Grenadier, Steven R., “Option Exercise Games: An Application to the Equilibrium Investment Strategies of Firms,” *Review of Financial Studies*, Summer 2002, **15**, 691–721.
- Karlin, Samuel, and Howard M. Taylor, *A First Course in Stochastic Processes, 2nd Ed.*, Academic Press, 1975.
- Kreps, David M, and Jose A. Scheinkman, “Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes,” *Bell Journal of Economics*, 1983, **14**, 326–37.
- Levenstein, Margaret C., and Valerie Y. Suslow, “What Determines Cartel Success?” *Journal of Economic Literature*, March 2006, **44**, 43–95.
- McAfee, R. Preston, Hugo M. Mialon, and Michael A. Williams, “What is a Barrier to Entry?” *American Economic Review*, May 2004, **94**.

- Murti, Pauli, "Exit in Duopoly under Uncertainty," *RAND Journal of Economics*, Spring 2004, **35**, 111–127.
- Official Journal of the European Union*, 5.2.2004, C31/12, February 2004.
- Novy-Marx, Robert, "An Equilibrium Model of Investment under Uncertainty," *Review of Financial Studies*, Sept. 2007a, **5**, 1461–1502.
- Novy-Marx, Robert, "Cournot Outcomes in Option Exercise Games," unpublished manuscript, Nov. 2007b.
- Pindyck, Robert S., "A Note on Competitive Investment under Uncertainty," *American Economic Review*, March 1993, **83**, 273–277.
- Pindyck, Robert S., "Sunk Costs and Real Options in Antitrust Analysis," in *Issues in Competition Law and Policy*, ABA Press, 2008
- Posner, Richard A., "The Chicago School of Antitrust Analysis," *University of Pennsylvania Law Review*, April 1979.
- Schmalensee, Richard, "Sunk Costs and Antitrust Barriers to Entry," *American Economic Review*, May 2004, **94**.
- Sutton, John, *Sunk Costs and Market Structure*, Cambridge: M.I.T. Press, 1991.
- U.S. Department of Justice and Federal Trade Commission, *Horizontal Merger Guidelines*, April 1997.
- Viscusi, W. Kip, John M. Vernon, and Joseph E. Harrington, Jr., *Economics of Regulation and Antitrust*, 2nd Ed., M.I.T. Press, 1995.

Table 1: Expected Number of Firms

σ_1	λ	$\mathcal{E}(n 1, 5)$	$\mathcal{E}(n 1, 10)$	$\mathcal{E}(\bar{n}_1)$
.2	0	1.39	1.64	1.43
.4	0	1.70	1.93	7.98
.2	.1	1.16	1.14	1.41
.4	.1	1.33	1.18	7.86

NOTE: $\mathcal{E}(n|1, 5)$ and $\mathcal{E}(n|1, 10)$ are the expected number of firms after 5 and 10 years, respectively, when there is initially one firm in the market. $\mathcal{E}(\bar{n}_1)$ is the expected value of the maximum number of firms that will ever be in the market, when there is initially one firm. Parameter values: $r = .04$, $\mu = .08$, $\alpha_1 = 0$.

Table 2: Dependence of $\bar{\omega}_n$ and S^*/S on Risk and Expected Return

	A No Restrictions on Parameters			B $\mu = r + .2\sigma_1$		C $\mu = r + .2\sigma_1$ $\alpha_1 = \alpha_{10} + \lambda$	
Parameter	$\bar{\omega}_5$	S^*/S	$\Delta S^*/S$	$\bar{\omega}_5$	S^*/S	$\bar{\omega}_5$	S^*/S
$\sigma_1 =$							
0	1.28	3.00	0	0.74	1.00	0.74	1.00
.2	1.78	5.80	4.29	1.78	5.80	1.78	5.80
.4	2.07	7.85	6.93	2.36	9.57	2.36	9.57
.6	2.58	12.23	10.13	3.00	15.56	3.00	15.56
.8	3.16	18.31	13.47	3.66	23.32	3.66	23.32
$\mu =$							
0.04	1.28	3.00	NA				
0.06	1.51	4.18	NA				
0.08	1.78	5.80	NA				
0.10	2.08	7.95	NA				
$\lambda =$							
0	1.78	5.80	4.29	1.78	5.80	1.78	5.80
0.10	2.55	11.90	4.36	2.55	11.90	1.68	7.88
0.20	3.19	18.68	4.59	3.19	18.68	1.92	12.30
$r =$							
0.02	1.91	13.37	10.09	1.59	10.74	1.59	10.74
0.04	1.78	5.80	4.29	1.78	5.80	1.78	5.80
0.06	1.69	3.50	2.52	1.95	4.12	1.95	4.12

NOTE: Each entry shows the value of $\bar{\omega}_n$ and S^*/S corresponding to the particular parameter values, for $n = 5$. Unless otherwise indicated, $r = .04$, $\mu = .08$, $\alpha_1 = 0$, $\sigma_1 = 0.2$, $\phi\rho_{\theta m} = 0.2$, and $\lambda = 0$. $\Delta S^*/S = (S_s^* - S_n^*)/S$ is the maximum differential impact of systematic risk.

Figure 1: Increase in Sunk Cost, S_2/S

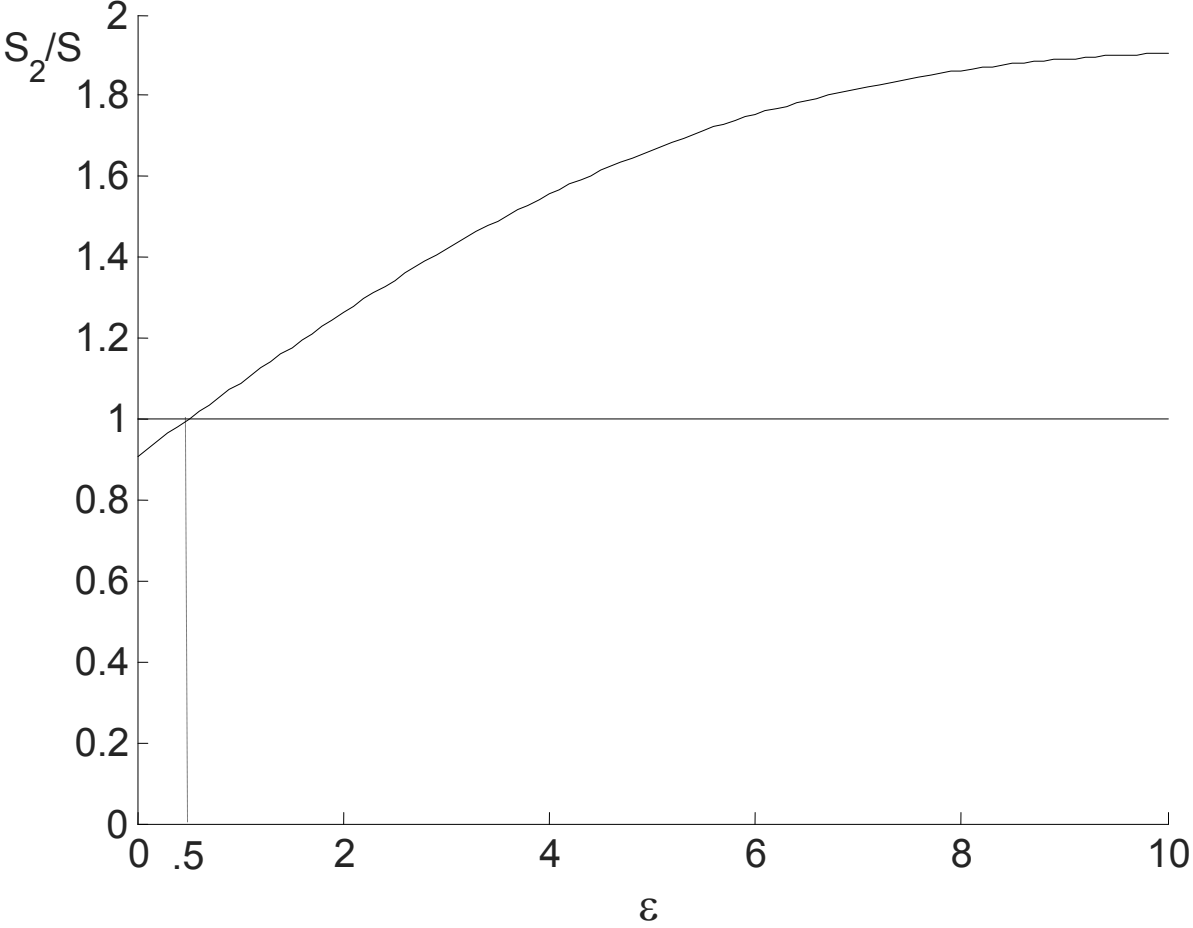


Figure 1: Increase in Sunk Cost

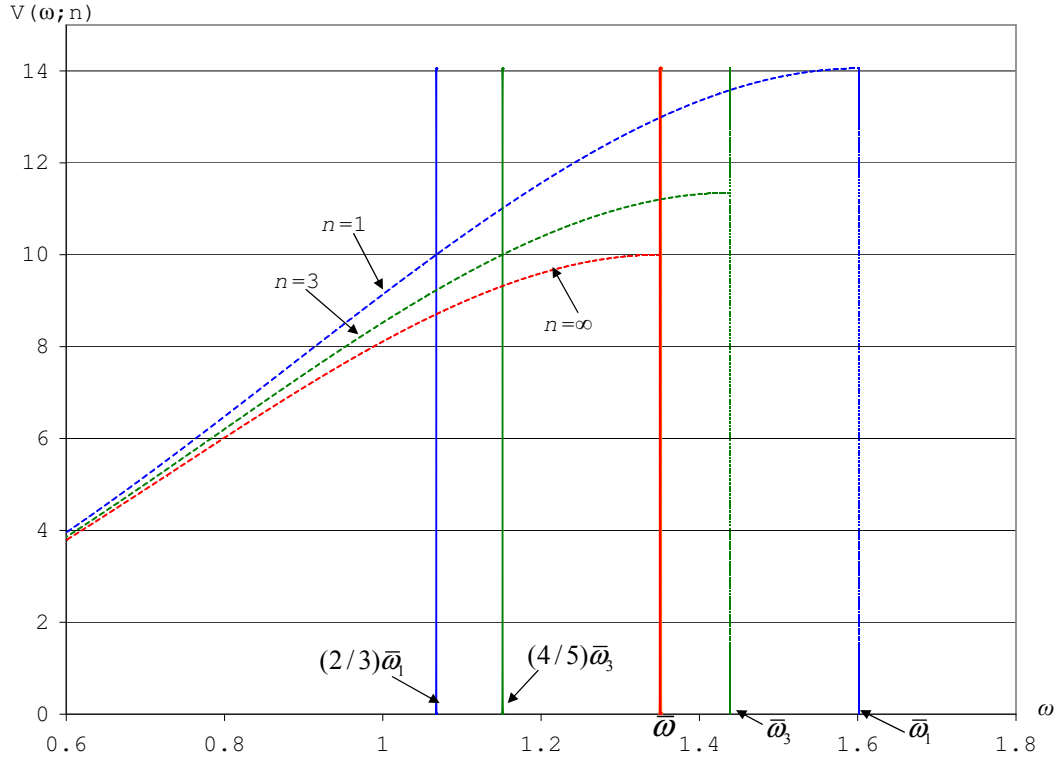


Figure 2: Firm Value and Entry Threshold
 Note: $r = .04$, $\mu = .08$, $\alpha_1 = \lambda = 0$, $\sigma_1 = .2$.

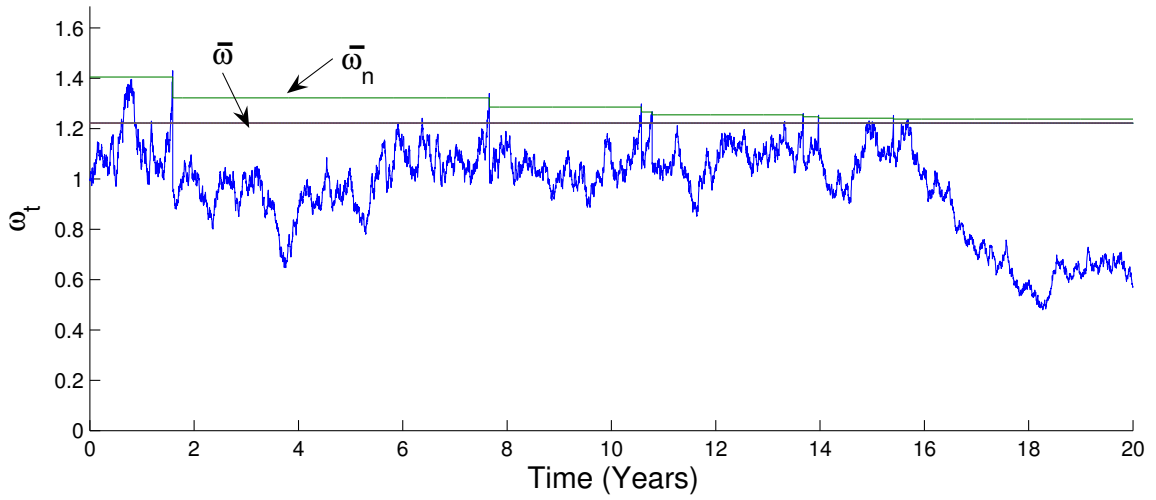
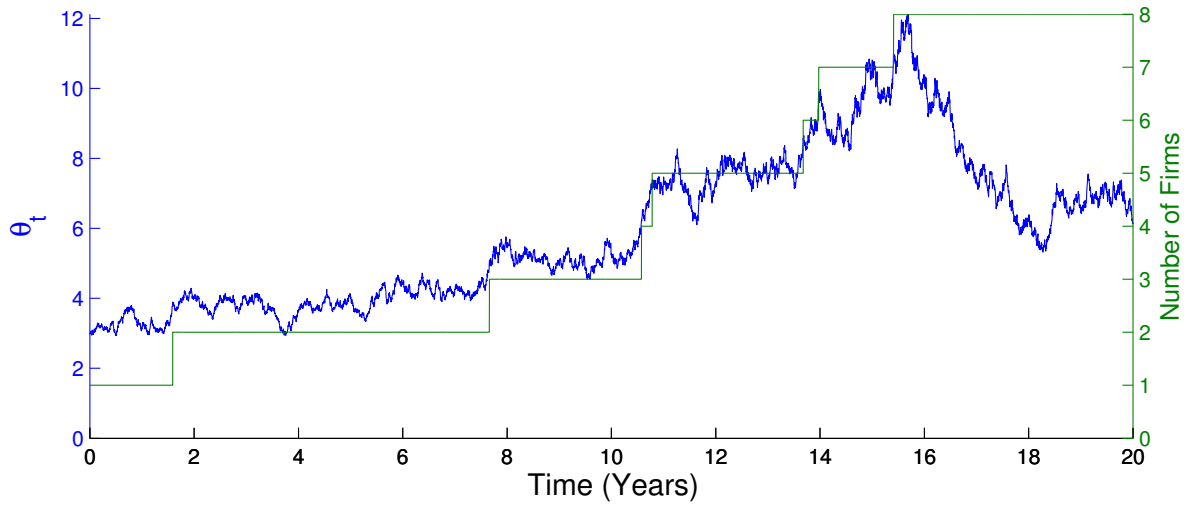


Figure 3: Sample Path: $\lambda = 0$

Note: $r = .04$, $\mu = .08$, $\alpha_1 = 0$, $\sigma_1 = .2$. Also, $\theta_0 = 3$, so that initially $n = 1$.

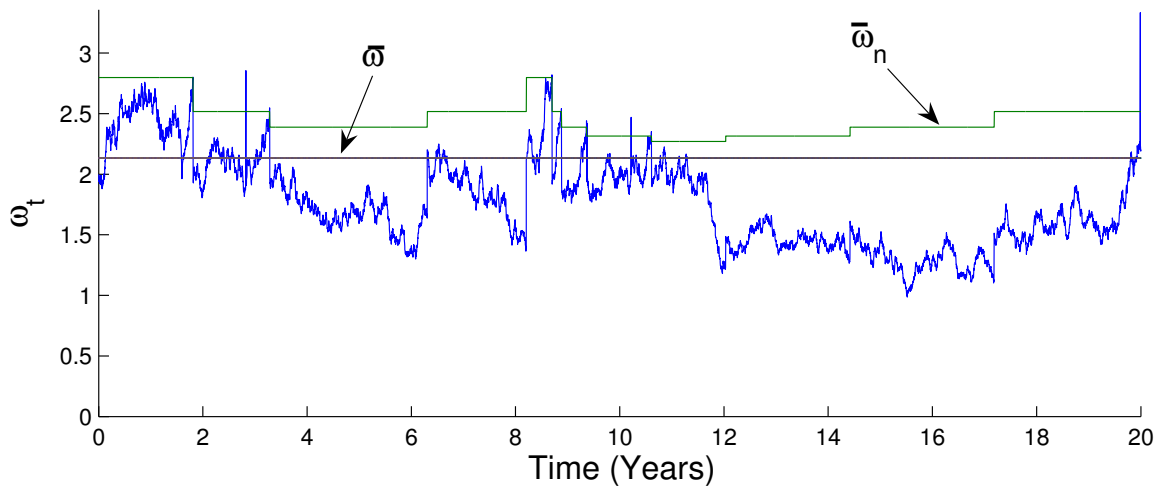
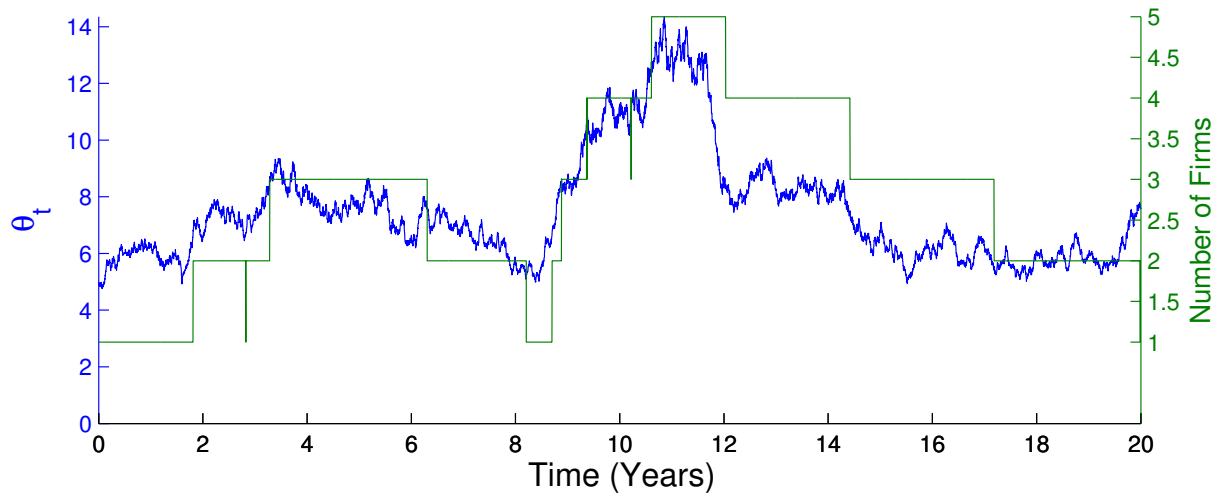


Figure 4: Sample Path: $\lambda = 0.1$

Note: $r = .04$, $\mu = .08$, $\alpha_1 = 0$, $\sigma_1 = .2$. Also, $\theta_0 = 5$, so that initially $n = 1$.