Essays on Competition,
Corporate Financing and Investment:
A Stochastic Dynamic Approach

PROEFSCHIFT

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg,
op gezag van de rector magnificus, prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen ten overstaan van een door het college voor
promoties aangewezen commissie in de aula van de Universiteit op vrijdag
24 oktober 2008 om 14.15 uur door

SEBASTIAN GRYGLEWICZ

geboren op 26 september 1977 te Walcz, Polen.

Contents

Acknowledgments

A number of people have made contributions—both direct and less direct—to this thesis. This note of acknowledgment must begin with an expression of my debt to my advisors, Kuno Huisman and Peter Kort. They have devoted a lot of time to me and have shaped my economic thinking. Our cooperation has been both stimulating and truly enjoyable. Kuno and Peter are co-authors of Chapter 4 of this thesis, but by numerous discussions and critical feedback they have significantly contributed to the other chapters as well.

I am thankful to the other Ph.D. committee members, Thomas Mariotti, Grzegorz Pawlina, Enrico Pennings, Han Smit and Dolf Talman for agreeing to join in and reading the manuscript. I am privileged to have them on the committee. I owe a special debt to Dolf Talman who gave me many detailed and valuable comments. I also thank Thomas Mariotti for hosting me during my visit at Toulouse University and for his feedback on my papers.

I thank Tilburg University, its staff and faculty members for providing me with a great environment to work on this thesis. In particular I would like to mention the assistance and guidance I received at various stages from Jan Boone, Eric van Damme, Wieland Mueller, Hans Schumacher and Thijs ten Raa. Many colleagues and friends at Tilburg University have contributed to this environment. I was fortunate to be affiliated with CentER at the same time as were Leon, Gala, Esen, Baris, Katya, Kim, Jeremie, Marta, Pedro, Miguel, Pedro, Mieszko, Viorel, Robert, Tunga, Andrea, Kenan Owen and others. I thank them for being a great companion on many academic and social occasions.

I owe a special gratitude to my friends who have positively, and typically indirectly, affected my creativity. Thanks to Marcin, Constanza, Tibor, Drahuska, Karin, Elif,

Barney and also my friends from Poland, who remained the best friends despite the geography, Marcin, Ula, Dawid, Szymon, Marcin, Blady, Przemek and Aneta.

I am most grateful to my parents for their continuing love, support and encouragement. I would like to use this opportunity to express my gratitude and to dedicate this thesis to them.

Finally, Agnieszka, my wife, has provided so much love and friendship that made the years of working on this thesis most enjoyable. I can only hope I will have contributed that much to her own thesis as she has to mine.

1

Introduction

This thesis studies the theory of dynamic economics and finance. The three important unifying principles underlying the presented analyses are the economic effects of time, uncertainty and information. These features are present in most economic situations, and this implies that decisions and interactions are influenced by intertemporal effects, uncertainty about the future, and information and learning.

More specifically, the thesis takes a direct approach to incorporate the following principles. First, agents make decisions continuously in time. For example, a firm that considers entering a new market has some flexibility with respect to the timing of the decision to enter. An incumbent firm in this market makes its pricing and capacity choices continuously in time, possibly taking into account the chances of entry. Second, the economic environments, for instance markets, are uncertain and evolve over time. In the same example, both entrant and incumbent firms face demand and productivity shocks that affect their entry and pricing strategies. Third, information is revealed in time and agents learn about the characteristics of the economic environment or other agents. While the entrant firm may initially have inferior information about some characteristics of the incumbent or the market, and so it may be incompletely informed about the profitability of entry, it may infer this information by observing through time the (pricing) strategy used by the incumbent. This specific example of entry and dynamic entry deterrence in a stochastic market is taken from Chapter 2 and indicates how time, uncertainty and information are closely related and how they interact even in simple setups. These three economic forces are central to many economic and financial situations and they are underlying the analyses presented throughout the thesis.

To analyze these effects, we base our theoretical analyses on continuous-time stochastic processes and their control. The dynamic stochastic setup allows us to build models

that are able to capture the complex roles of time, uncertainty and information. Furthermore, the continuous-time framework—and the use of stochastic calculus—makes the mathematics of the models tractable. Most importantly, our modeling choice delivers tools for advance study of relevant economics situations.

The thesis is a collection of four research papers. Each paper addresses an open economic problem from a theoretical perspective.

Chapter 2 contributes to dynamic game theory and, with an application of the general model, to industrial organization theory. We study a dynamic signaling game played in a stochastic environment. The standard signaling game is a two-player game of incomplete information, in which one player (uninformed) does not observe directly the type of the other player (informed). The type is a payoff-relevant characteristic and can be inferred from the actions (signals) chosen by the informed player. This setup has been one of the most popular games in applications in industrial organization, corporate finance and labor economics. While these disciplines have often benefited from more advanced analyses in multi-period models under uncertainty, the models of signaling situations have, in general, remained in one-period simple setups. Yet, the interactions between agents are most often repeated and take place in evolving environments. The aim of Chapter 2 is to analyze these additional effects and provide a framework to study multi-period stochastic signaling games.

Specifically, the model presented in Chapter 2 introduces a class of two-player signaling games in continuous time in which the stake contested by the uninformed player is a diffusion process observed by both players. We suppose that the payment of the stake depends on the privately-observed type of the informed player and that the informed player of one type can, at cost, imitate other types. We show that the signaling game is played as long as the stake stays within two-sided bounds on the state variable (stake). In equilibrium the informed player reveals her type at a randomized lower trigger. The uninformed player learns about the true type by observing the minimum process of the stake and contests the stake at an upper boundary that is decreasing in the running minimum.

We then apply the game to model dynamic limit pricing under stochastic demand and derive a set of inferences unavailable in one-period deterministic models. The limit pricing model based on incomplete information was introduced by Milgrom and Roberts (1982) and it studies an incumbent firm that uses prices as an instrument to signal unprofitable entry and deter a potential entrant. We adapt our general signaling model to a limit pricing problem by interpreting the diffusion process as stochastic demand, the informed player as the incumbent and the uninformed player as the entrant. One advantage of our dynamic setup is that it generates equilibrium price dynamics and, specifically, that price dynamics may reveal limit pricing of incumbents.

In equilibrium the limit-pricing incumbent reveals its type by increasing prices as the market becomes unfavorable to entry. This means that increasing prices in a decreasing market may indicate entry deterring limit pricing. The model also implies that, despite that the demand is modeled as a Markovian variable, the decision to enter exhibits path dependence and the entrant assessment of entry profitability depends not only on the current state of the market, but also on the historical minimum.

Chapter 3 is a contribution to corporate finance theory. It analyzes the effects of financial distress on corporate financing choice and other financial decisions. In contrast to the existing literature, we study both short-term liquidity and long-term solvency concerns. From the modeling point of view, our contribution can be seen as incorporating two strands of literature in an analytically tractable framework. One strand of literature originates from the contingent claims models of risky debt of Black and Scholes (1973) and Merton (1974), and is developed in a popular trade-off framework of corporate finance by Leland (1994). These models have been successful in studying capital structure choice, solvency default and credit risk, but have failed to incorporate corporate liquidity risk, realistic dividend policy and cash holdings. The other strand of literature, represented by Jeanblanc-Picqué and Shiryaev (1995), studies dynamic dividend payout optimization with liquidity shocks. These models typically lack financing choice and solvency concerns. And, remarkably, they have failed to produce a model of smooth dividends, which is one of the most pervasive characteristics of corporate dividends (Lintner (1956), Brav, Graham, Harvey and Michaely (2005)).

To put the model of Chapter 3 in the context of time, uncertainty and information, consider a firm that seeks financing from a combination of debt and equity. Once financed, the firm generates an uncertain stream of cash flows. At each time the firm divides net profits or losses into dividend payments to equity and retained earnings to increase cash holdings. Negative cash shock can lead to default: either solvency default, if the firm is not profitable enough, or liquidity default, if the firm has no liquidity to cover its debt obligations. Both liquidity shocks and profitability level are uncertain, and the firm learns the true long-term profitability by observing the realizations of cash flows.

Extending the contingent claims trade-off model with liquidity concerns offers a wide range of implications for corporate finance. We show that there are important interactions between liquidity and solvency. Since a less solvent firm requires less cash to cushion liquidity shocks before becoming insolvent, lower solvency results in higher corporate liquidity. On the other hand, because raising cash to cover liquidity requirements is costly, liquidity affects financing decisions, and, via optimal capital structure choice, corporate solvency. The model provides a rationale for significant corporate cash reserves and produces a dynamic cash policy that is in line with empirical regularities.

Because of the interplay of liquidity and solvency concerns, positive cash flow shocks are retained and negative shocks decrease the optimal cash reserves. Consequently, the optimal dividend distributions are smoothed relative to cash flows. The introduction of liquidity concerns addresses some of the critiques towards the predictive power of structural models. First, in an empirical study Eom, Helwege and Huang (2004) report that the common problem of structural models is that the predicted spreads are too dispersed. Our model predicts a lower dispersion of credit spreads across firms than the model without liquidity. Second, the standard structural models tend to predict too high leverage ratios. By including cash reserves, our analysis predicts a significantly lower share of debt in firm value.

Chapters 4 and 5 study corporate investment decisions. The recent literature on investment has stressed three characteristics that hold for most investment decisions. First, a firm cannot costlessly adjust its capital stock, i.e., investment is irreversible or partially irreversible and involves some sunk cost. Second, future cash flows are uncertain. Third, firms in general have some flexibility with respect to investment timing. Investment projects with these characteristics can be seen as options (opportunities without obligations) to invest and investment decisions are timing decisions about when to exercise these options. To express the analogy to financial options, the now-prevailing approach to (real) investment analysis is called the real options approach. Dixit and Pindyck (1994) provide an introduction and review of early contributions.

The option-based approach in modeling investment has a significant effect on optimal investment decisions. The standard approach based on the net present value (NPV) rule prescribes that an investment should be undertaken whenever discounted expected future revenue flows exceed current outlays, implying that the NPV is positive. However, this rule does not take into account the loss of flexibility at the time of the investment. The loss arises, because by investing the firm gives up the opportunity to wait for new information and to decide at a later stage whether to invest or refrain from investment. This opportunity is called the value of waiting and it must be included as one of the costs of investment. The above characterization of the investment problem clearly encompasses the three recurring aspects of this thesis, namely, time, uncertainty and information. In Chapters 4 and 5 we contribute to the real options literature by studying two novel investment (or divestment) problems.

Chapter 4 revisits the important result of the real options approach to investment, which states that increased uncertainty raises the value of waiting and thus decelerates investment. Typically, in this literature projects are assumed to be perpetual. However, in today's economy firms face a fast-changing technology environment, implying that investment projects are usually considered to have a finite life. Our analysis in this chapter studies investment projects with finite project life, and we find that, in contrast

with the existing theory, investments may be accelerated by increased uncertainty. It is shown that this particularly happens at low levels of uncertainty and when project life is short. Chapter 4 is based on Gryglewicz, Huisman and Kort (2008).

Chapter 5 studies optimal divestment policy of a firm that may partially and gradually divest its capital or sell the whole firm at once. Partial divestment offers greater flexibility while a whole-firm transaction provides a price premium. We show that, if the price premium includes both a fixed and a proportional component, a large firm optimally starts to divest partial capital before choosing to sell the whole-firm. It turns out that full-firm divestment is preferable over partial divestment with higher profit volatility, in more declining markets and if capital is less industry-specific.

The thesis also has its methodological contributions. The techniques of stochastic control are used in innovative ways to solve novel economic problems. Chapter 2 applies the theory of optimal control of extremum processes to study learning about unknown types of other players. We start with formulating the problem with two Markov state variables, that is the payoff of the game and Bayesian belief about the other player's type. We show that the original problem with complicated Bayesian updating can be translated in a substantially simpler problem, in which the belief state variable is replaced by the minimum process of the payoff variable. We can then use the very tractable framework of optimal stopping of maximum processes (see Peskir (1998) and Peskir and Shiryaev (2006)) to solve the problem of the uninformed player.

In Chapter 3 we introduce unknown drift and filtering to model two sources of uncertainty, namely short-term liquidity shocks and uncertain long-term solvency. Short-term uncertainty is represented directly by unpredictable Brownian increments of cash flows. To capture long-term uncertainty, we assume that the value of mean instantaneous cash flow is initially uncertain, but has a known distribution, and the realizations of the stochastic process are used to learn about the true nature of the cash flow process. This characterization has the desirable feature that persistent liquidity shocks translate into solvency shocks (for example, persistent negative liquidity shocks indicate low profitability). The filtering formulation of cash flow dynamics allows us to develop a model of corporate finance that is parsimonious and analytically tractable, yet broad in scope and rich in predictions.

In Chapter 5 we apply a combination of barrier control and optimal stopping to analyze the costs and benefits of marginal versus discrete adjustments of capital. Marginal adjustments of capital, modeled as a barrier control problem, leave the firm with greater flexibility. This flexibility is valuable in stochastic environments and remains so even if capital adjustments are irreversible. On the other hand, irreversible discrete capital investment (or divestment), modeled as an optimal stopping problem, is less

flexible, but is frequently attributed with a price premium. The combination of barrier and optimal stopping control is new in the context of real options analysis.

2

A Stochastic Version of the Signaling Game

2.1   Introduction

Many important economic situations that involve incomplete information and signaling are set in dynamic stochastic environments. A monopolist that uses prices to signal unprofitable entry, as in the limit pricing model of Milgrom and Roberts (1982), in reality has to do so repeatedly and under changing market conditions. A firm that uses dividends to signal its profitability, as in e.g. Miller and Rock (1985), typically makes payout decisions repeatedly while facing stochastic profit flows. Yet, the available models based on signaling games allow only for one-time signals or repeated signals under stationary conditions. The purpose of this chapter is to extend the signaling game to a fully dynamic stochastic model.

We study a new class of two-player signaling games in continuous time in which the stake contested by the uninformed player is a diffusion process X_t observed by both players. The informed player's type is either strong or weak and initially her type is only privately observed. The uninformed player obtains the stake if he contests it from the weak type, but receives nothing if the other player is strong. The informed player of the weak type gets a negative payoff if contested, while the strong type is unaffected by the contestant. It is possible, but costly, for the informed player to send signals and keep the other player uninformed about her own true strength.

How does the informed player signal in such a setting? How does the uninformed player make the strategic decision to contest the stake? Our primary insight is that, as the stake evolves in a stochastic environment, at some point in time the incentive constraints may stop being binding. In particular, provided that the game starts at a pooling situation, the uninformed player wants to contest as the stake gets high

enough given his belief about the other player's type. On the other hand, if the stake gets low enough, there is little threat from the contestant, so the informed player of the weak type does not want to send costly signals anymore and, consequently, prefers to reveal her type.

The decisions of the two players are strategically interrelated. The best responses are optimal stopping decisions, with the uninformed (respectively informed) player seeking a critical high state U (respectively low state L) to stop the signaling game. The equilibrium strategy U of the uninformed player balances the benefits of winning the stake with some current belief that the type is weak and the cost of loosing the opportunity to learn the revealed type if the stake reaches L. The strategy L of the weak informed player strikes the balance between the benefits of facing the uninformed contestant as late as U but bearing the signaling cost and the benefits of revealing its type, and not paying the signaling cost but facing early entry of the contestant that is sure to win the stake.

We show that the stopping game in general has no Markov perfect equilibrium in pure strategies. Instead, in equilibrium the informed player reveals her type at a randomized lower trigger. The reason for this is that a deviation from a pure strategy L provides a discrete gain while it bears an infinitesimal cost. Specifically, suppose that L is an equilibrium pure strategy of the weak player. If no action is taken at L, the uninformed player updates its belief and is certain to face the strong informed player. Then a slight deviation from L discretely improves the standing of the informed player against the uninformed one, while the cost of additional signaling is, due to continuous time, infinitesimal. A similar effect that information-revealing actions are played in a mixed strategy is also present in Huddart, Hughes and Levine (2001) in a context of informed trading.

The mixed type-revelation strategy introduces a remarkable Bayesian learning process based on the path of the stake X. The uninformed player, observing no actions in the support of the strategy of the informed player, updates his belief about the true type. Therefore the minimum process $M_t = \min_{0 \leq s \leq t} X_s$ can be used as a state variable governing the belief process. Based on these observations, the uninformed player's problem is non-trivial and involves path-dependent payoffs and learning from the path of the diffusion process. To solve it we use some recent developments in the theory of optimal control of extremum processes. Our characterization of the Markov perfect Bayesian equilibrium is relatively basic and is a solution to two ordinary differential equations subject to boundary conditions. In particular we show that in equilibrium the weak informed player reveals at a mixed-strategy lower trigger that is continuously distributed over some interval in X. The uninformed player contests the stake at an upper boundary that is decreasing in the running minimum on the same interval.

Two extensions from the standard one-period (two-dates) signaling models generate the interesting strategic interactions with learning in our model. These are multi-period dynamics and uncertainty about the future stake. In a one-shot game with a stochastic stake, there is no room for waiting to obtain information in the future. On the other hand, in a multi-period game with a fixed stake (or varying in time but deterministic and monotone), the strategic situation is non-trivial only at the initial node. Beyond the initial date, no learning and no type revelation can happen. Consequently, similar results to the ones presented in this chapter can be obtained in other setups that have multi-period signaling in stochastic environment. We choose for the continuous time framework which is standard and tractable in studies of timing and stopping games (e.g. Fudenberg and Tirole (1986), Bulow and Klemperer (1999), Dutta and Rustichini (1993)). To incorporate a stochastic environment in a tractable way we model the stake as a geometric Brownian motion.

By introducing continuous-time dynamics and uncertainty, our model is able to provide new insights into some of the well-known signaling situations in economics. As an illustration we apply the generic model to entry deterrence by limit pricing. We can translate our setting in a limit pricing problem by interpreting the diffusion process as stochastic demand, the informed player as the incumbent firm and the uninformed as a potential entrant. One advantage of our setup is that we can explore equilibrium price dynamics. The stochastic limit pricing game implies that price dynamics may reveal limit pricing of incumbents. Specifically, in equilibrium the limit-pricing incumbent reveals its type by increasing prices as the market conditions get unfavorable to entry. To an external observer this means that increasing prices in a decreasing market may be interpreted as an indicator of entry deterring limit pricing. This observation brings forward a policy instrument to detect anti-competitive pricing practices. This is in contrast to the standard one-shot signaling models of limit pricing that provide little in terms of antitrust policy recommendations. We also show that the decision of the entrant to enter exhibits path dependence. Specifically, despite the fact that the demand is modeled as a Markovian variable, the entrant assessment of entry profitability depends not only on the current market, but also on the historical minimum.

A few previous studies consider dynamic aspects of signaling. Saloner (1984) presents a multi-period version of the limit-pricing model of Matthews and Mirman (1983) in which signals received by the uniformed firm are noisy. In contrast, we assume that actions of the informed player are observed directly by the uniformed party. The key difference of our model is that we allow for a stochastic environment that changes over time. This means that there are states when signaling is more profitable and times when it is not profitable. In the paper of Saloner (1984), demand is uncertain, but demand shocks, that last for a single period, serve solely as a device to add noise

to the informed player's actions. The market conditions for both payers are identical before each round. Mester (1992) analyze a three-period signaling setting in which the unobservable type changes over time. Toxvaerd (2007) adapts and extends a similar setup to study limit pricing. Instead, we assume that the *observable* market conditions fluctuate but the unobservable type of the informed player is fixed. In our setting we analyze richer dynamics of the stochastic variable and effects of good and bad states on signaling strategies. Kaya (2007) studies separating equilibria in an infinitely repeated discrete-time signaling game, while we concentrate on how pooling and semi-separating equilibria can be sustained in a stochastic environment. Additionally, none of these models are formulated in continuous time.

The next section sets up the model. In Section 2.3 we study special cases regarding complete information and a deterministic environment. Section 2.4 presents the equilibrium analysis in the stochastic model. In Section 2.5 we apply the general signaling model to analyze limit pricing under stochastic demand. Section 2.6 concludes, and an Appendix collects the proofs omitted in the main text.

2.2   Model

2.2.1   Setup

The game is set in continuous time with infinite horizon, indexed with $t \in [0, \infty)$. There are two players. Player 1 is of type $\theta \in \{w, s\}$ (weak or strong) and knows her type. If Player 1 does not make any effort, θ is observed by Player 2. By exercising some costly effort, Player 1 of the weak type can mimic the behavior or appearance of the strong type and in this way pretend to be of type s. The cost of imitation per unit of time is $c > 0$. Player 2 has a prior belief $\pi_0 \in (0, 1)$ that $\theta = w$. Observing Players 1's actions, Player 2 updates his belief about θ using Bayes rule whenever possible and the belief at time t is denoted by π_t. In particular, at the first time the w-type ceases to send the signal, the belief is updated to 1.

Player 2 contests the stake of the game which exogenously evolves over time according to a geometric Brownian motion

$$dX_t = \mu X_t dt + \sigma X_t dZ_t,$$

with $X_0 = x_0$. The constants μ and $\sigma > 0$ are drift and volatility parameters. Z_t is a standard Brownian motion. At any (stopping) time $t \geq 0$, Player 2 can contest the stake by paying a fee of $K > 0$ (K could be interpreted as an entry cost or a checking fee; throughout the chapter we refer to Player 2 contesting the stake also as entering or stopping). At this time, Player 2 learns θ and gets a payoff X_t if $\theta = w$ and a

zero payoff if $\theta = s$. When checked by Player 2, Player 1 gets a fixed negative payoff normalized to -1 if she is of type w and 0 if she is of type s.

Both players are risk neutral and discount flows at a constant discount rate r. To concentrate on the interesting cases we assume throughout the chapter that $c < r$, which is a necessary condition for Player 1 of the w type to have an incentive to use signals to postpone entry. Similarly, we assume $\mu < r$ to guarantee convergence of the problem of Player 2.

Some aspects of our modeling strategy deserve comment. The choice of the particular payoff functions stems from our objective to keep the analysis simple and to incorporate the following desired features of the game. The environment X is stochastic and the payoffs depend on the state of the environment. In particular, the uninformed player wants to contest if the state is 'good' (high X). The uninformed player is worse off against the strong type of the informed player. Absent the cost of signaling, the weak type prefers to be recognized as a strong one. Finally, signaling is costly. We do not aim here to show the most general functional forms that support our results. Certainly our analysis can accommodate other payoff structures that preserve the above mentioned features. Indeed, we consider an example of limit pricing in Section 2.5 where payoffs are in flows and the signaling cost and the informed player's payoff depend directly on the state X.

For clarity of the exposition, the model has the signaling part of the game in a reduced form. We do not explicitly model the signals used by the w type. The implicit assumption is that the pooling strategy in equilibrium is the efficient one, i.e. the one least costly to Player 1 (the interpretation is particularly straightforward in the binary case mentioned in the next paragraph).

A further simplifying assumption is that the s type is infinitely strong and is indifferent to entry. Thus we avoid discussing separating strategies. Our analysis conveys to the case in which the s type has little incentives to separate (i.e. the s type looses little when contested or when the cost of separation is high) and therefore does not separate (we indeed assume that the s type is less than infinitely strong in Section 2.4 to prove uniqueness of equilibrium strategies). Another important situation to which our model applies directly is when the signaling space is limited. The sharpest, yet frequently realistic, case is the binary signaling space which leaves little room for separating strategies. For concreteness one could interpret our baseline model as a game with a binary signaling space. Finally, we note that in any case a separating equilibrium would be less remarkable in our setup as it would eliminate the interesting belief dynamics.
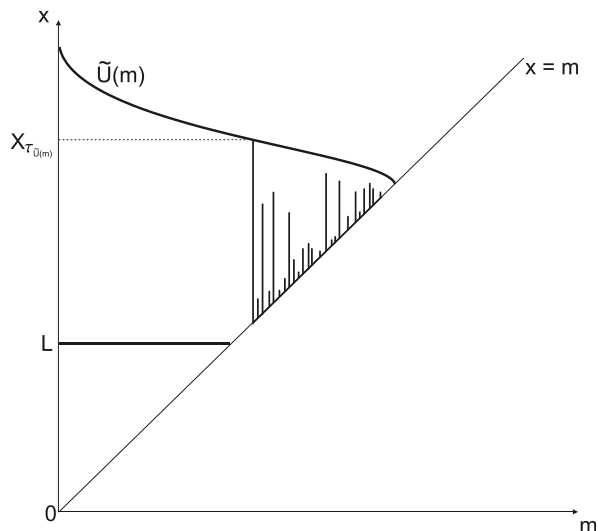
FIGURE 2.1. A sample path of (x, m).

2.2.2   State space, strategies and Markov equilibria

In principle, the state space of payoff relevant variables consists of a pair of Markov state variables $(x, \pi) \in \mathbb{R}_+ \times [0, 1]$. It will simplify our analysis if we transform the state space based on the following important observation. As we will show in Section 2.4.2, in the continuation region, before the game is stopped, the belief variable π_t is a function of the running minimum of X_t, that is of $M_t = \min_{0 \le s \le t} X_s$. The intuition is that, as Player 1 may prefer to reveal her type as the stake becomes low, the historical minimum may be used by Player 2 to update his belief about θ. At most parts of the analysis, it will be more convenient to work with the minimum than with the beliefs so, where indicated, we analyze the game in the state space $\left\{ (x, m) \in \mathbb{R}_+^2 : x \ge m \right\}$.

Player 1 of the weak type takes an action $a_1 \in \{signal, reveal\}$, where $a_1 = reveal$ indicates a decision to stop signaling and to reveal the player's type. As the strong type is passive, in the sequel we shall implicitly mean the weak type, when we discuss actions and strategies of Player 1. Once the type is revealed, the game becomes a game of complete information with no strategic interactions. Given optimal behavior of the players beyond this point, the expected discounted payoffs are considered as the terminal payoffs of the signaling game. To specify these termination payoffs suppose that the w type reveals at time t and denote by $\tau_w \ge t$ the (stopping) time at which Player 2 would optimally collect the stake upon the payment of K. Then the expected discounted payoff at time t of Player 2 is $e^{-r(\tau_w - t)} \left(X_{\tau_w} - K \right)$ and of Player 1 is $-e^{-r(\tau_w - t)}$ (the details of the analysis for the complete information case are carried out in Section 2.3.1). Player 2 takes an action $a_2 \in \{do\ not\ contest, contest\}$. Once $a_2 = contest$ is chosen, the signaling game is over and the terminal payoffs are collected.

Threshold		$x = L(\pi)$	$x = U(\pi)$
Stopping time		τ_L	τ_U
Player 1	w type	$-e^{-r(\tau_w - \tau_L)}$	-1
	s type	$-$	0
Player 2		$e^{-r(\tau_w - \tau_L)}(U(1) - K)$	$\pi_{\tau_U} U(\pi_{\tau_U}) - K$

TABLE 2.1. Terminal payoffs.

Markov strategies prescribe actions to the current state. In the signaling game, Markov strategies define two sets in the state space, a continuation set ($a_1 = signal$ and $a_2 = do\ not\ contest$) and a stopping set ($a_1 = reveal$ and $a_2 = contest$). In other words, each player faces an optimal stopping problem. The theory of optimal stopping of Markov processes indicates that the strategies will take the form of an optimal stopping boundary (see Peskir and Shiryaev (2006)). In the case of Player 2, define $U : [0, 1] \to \mathbb{R}_+$, then $U(\pi)$ is a *boundary* separating continuation and stopping regions in the state space (x, π). Precisely, Player 2 chooses '*do not contest*' if $x < U(\pi)$ and chooses '*contest*' whenever $x \geq U(\pi)$. The associated stopping time is defined as $\tau_U = \inf \{\tau \geq 0 : X_\tau \geq U(\pi)\}$. Analogously, if the strategy of Player 2 is considered in the state space (x, m), the function $\tilde{U} : \mathbb{R}_+ \to \mathbb{R}_+$ is a free boundary such that Player 2 '*does not contest*' if $x < \tilde{U}(m)$ and '*contests*' whenever $x \geq \tilde{U}(m)$.

A solution to the problem of Player 1 takes the form of a lower boundary $L(\pi)$. Player 1 signals as long as $x > L(\pi)$ and reveals as soon as $x \leq L(\pi)$. Formally, the strategy prescribes a stopping time $\tau_L = \inf \{\tau \geq 0 : X_\tau \leq L(\pi)\}$. The stopping time at which Player 2 enters after the type is revealed denoted above as τ_w is equal to $\inf \{\tau \geq \tau_L : X_\tau \geq U(1)\}$.

As an illustration of the two-dimensional process (x, m), Figure 2.1 presents a sample path in the continuation region between some L (here constant) and $\tilde{U}(m)$ (here decreasing in m) with a realization of the stopping rule at X_{τ_U}. Above the diagonal $x = m$, the process evolves vertically reflecting changes in x. When new minima are reached, the process moves down along the diagonal. Since the process stays above L and reaches first $\tilde{U}(m)$, here Player 2 contests before Player 1 has revealed her type.

Note that Figure 2.1, which aims to present the (x, m) process, does not necessarily represent any equilibrium or sensible strategies. In fact in a game in pure strategies, starting at π_0, the belief can be only updated to either 0 or 1 and the signaling game is stopped at either $L(\pi_0)$ or $U(\pi_0)$ and thus $\tilde{U}(m)$ is constant for $m > L(\pi_0)$.

Table 2.1 collects the information about the terminal payoffs when the signaling game is stopped at either $L(\pi)$ or $U(\pi)$. Given a pair of (pure) strategies (L, U) and the starting values of x and π, the respective total expected payoffs of Player 1 and 2

are given by

$$
\begin{aligned}
R_1(x, \pi; L, U) &= \mathbb{E}_{x,\pi} \left[- \int_0^{\tau_L \wedge \tau_U} e^{-rt} c\, dt - e^{-r\tau_U} 1_{\tau_L \geq \tau_U} - e^{-r\tau_w} 1_{\tau_L < \tau_U} \right], \\
R_2(x, \pi; L, U) &= \mathbb{E}_{x,\pi} \left[e^{-r\tau_U} \left(\pi_{\tau_U} U(\pi_{\tau_U}) - K \right) 1_{\tau_L \geq \tau_U} + e^{-r\tau_w} (U(1) - K) 1_{\tau_L < \tau_U} \right].
\end{aligned}
$$

It is essential for our analysis to allow the players to randomize across pure strategies. As we show later, in the general case the game has no perfect equilibrium in pure strategies. As our subsequent analysis focuses on the case in which Player 1 applies a mixed strategy and Player 2 responds with a pure strategy, we need to consider only a mixed strategy of Player 1. A *mixed strategy* of Player 1 is a probability measure P on $[0, x_0]$ with the corresponding distribution function G defined by $G(x) = P([x, x_0])$.[1] G is interpreted as a distribution function over trigger strategies. The expected payoff that corresponds to a pair of strategies (G, U) is $R_i(x, \pi; G, U) = \int R_i(x, \pi; L, U) dG$ for $i = 1, 2$.

A *Markov perfect Bayesian equilibrium* (MPBE) is a pair of Markov strategies (G^*, U^*) such that

$$
\begin{aligned}
R_1(x, \pi; G^*, U^*) &\geq R_1(x, \pi; G, U^*), \\
R_2(x, \pi; G^*, U^*) &\geq R_2(x, \pi; G^*, U),
\end{aligned}
$$

for all states (x, π) and all strategies G and U.

2.3   Simple cases

2.3.1   Complete information

Suppose that for some $t \geq 0$ $\pi_t \in \{0, 1\}$, so that signaling does not play a role in the game. If $\pi_t = 0$, i.e. Player 1 is strong with probability one, then obviously Player 2 never tries to contest the stake, i.e. $U(0) = \infty$. If $\pi_t = 1$, then Player 2 solves the optimal stopping problem

$$
W(x) = \sup_{t \leq \tau \leq \infty} \mathbb{E} \left[e^{-r(\tau - t)} \left(X_\tau - K \right) \middle| X_t = x \right].
$$

As usual, the optimal strategy in the solution of the problem takes the form of an upper trigger. Let $U(1)$ denote the stopping threshold. By the standard dynamic programming argument and Itô's lemma, $W(x)$ satisfies the following Hamilton-Jacobi-Bellman

---

[1] Our definition follows the logic of mixed strategy, i.e. a player chooses a randomized pure strategy, which in our case is a stopping time. More detailed discussion of mixed and behavioral strategies and their equivalence in stopping games can be found in Touzi and Vieille (2002).

differential equation

$$rW(x) = \mu x W'(x) + \frac{1}{2}\sigma^2 x^2 W''(x) \tag{2.1}$$

in the continuation region, i.e. for $x \in (0, U(1))$. The left-hand side (2.1) reflects the required rate of return per unit of time for holding the option to get x. The right-hand side is the expected change in the value of the option.

The differential equation is associated with the following three boundary conditions:

$$W(U(1)) = U(1) - K, \tag{2.2}$$
$$W'(U(1)) = 1, \tag{2.3}$$
$$W(0) = 0. \tag{2.4}$$

The value matching (2.2) and smooth pasting (2.3) conditions impose a continuous and smooth fit at the boundary, required for optimality. Condition (2.4) ensures that the stake will be worthless if x reaches its absorbing barrier zero. Solving equations (2.1)-(2.4) we obtain that Player 2 optimally contests the stake at the stopping time

$$\tau_w = \inf\{\tau \geq t : X_\tau \geq U(1)\},$$

where

$$U(1) = \frac{\beta_1 K}{\beta_1 - 1}, \tag{2.5}$$

and

$$\beta_1 = \frac{1}{2} - \frac{\mu}{\sigma^2} + \sqrt{\left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}}.$$

β_1 is the positive root of the characteristic equation $\frac{1}{2}\sigma^2\beta(\beta - 1) + \mu\beta - r = 0$ and is always larger than 1. The value $W(x)$ is given by[2]

$$W(x) = \left(\frac{x}{U(1)}\right)^{\beta_1}(U(1) - K). \tag{2.6}$$

Knowing the optimal behavior of Player 2, let Ω denote the expected discounted value of the weak type of Player 1 if the type is revealed at time t. Then $\Omega(x) = E\left[-e^{-r(\tau_w - t)}|X_t = x\right]$. Using similar methods as above we obtain that

$$\Omega(x) = -\left(\frac{x}{U(1)}\right)^{\beta_1}. \tag{2.7}$$

---

[2] The intuition for this and some other similar expressions in this chapter can be gained by observing that $E[e^{-r(\tau - t)}|X_t = x] = (x/X_\tau)^{\beta_1}$, where τ is a stopping time. Thus, given that the current state is x, $(x/X_\tau)^{\beta_1}$ is the present value of one dollar received at stopping time τ.

$\Omega(L)$ and $W(L)$ are the respective terminal payoffs if Player 1's type is revealed at threshold L (see column $x = L$ in Table 1).

2.3.2   Deterministic X_t

Suppose now that $\sigma = 0$, so that the process $dX_t = \mu X_t dt$ is deterministic. In this case, the strategies in the unique equilibrium differ depending on the sign of μ, but they share a similar simple structure. For any given belief π, Player 2 contests the stake as he would do in a non-strategic situation. Given this behavior of Player 2, Player 1 applies a straightforward incentive constraint to decide between signaling or revealing (or randomizing between these two).

We first analyze the simpler case in which x decreases in time. If $\mu < 0$ then, for any fixed π, Player 2 would not postpone the entry decision and would rather enter immediately, if at all. If x falls below K, Player 2 never enters and so Player 1 does not signal.

Proposition 2.1 *If $\sigma = 0$ and $\mu < 0$, the signaling game has a unique MPBE. In the equilibrium Player 2 applies the following upper boundary strategy*

$$U(\pi) = \frac{K}{\pi}.$$

The lower trigger strategy of Player 1 of the w type is

$$L = \begin{cases} x_0 & if \ \Gamma_1(x_0) \leq 0, \\ K & if \ \Gamma_1(x_0) > 0, \end{cases}$$

where

$$\Gamma_1(x) = -\frac{c}{r} + \left(\frac{x}{K}\right)^{\frac{r}{\mu}} \frac{c}{r} + 1.$$

$\Gamma_1(x_0)$ characterizes the incentive compatibility constraint of Player 1 for signaling from $t = 0$ until the time when x reaches K. (Note that the weak and strict inequalities in the strategy L are rather arbitrary, but the events with equality are of measure zero.) It is simply the difference between the total cost of signaling and revealing its type immediately in the situation that this would trigger immediate entry of Player 2.

If $\mu > 0$ and x increases in time, several additional considerations arise. First, for any $\pi > 0$ there is a sufficiently large x such that Player 2 decides to contest the stake. Second, apart from the full separation and pooling outcomes arising also in the case of negative μ, if $\mu > 0$ there is an intermediate set of parameters at which the unique outcome is semi-separation. Third, agents take into account the value of waiting (similar to the stochastic, complete information case of Section 2.3.1).

Proposition 2.2 *If $\sigma = 0$ and $\mu > 0$, the signaling game has a unique MPBE. In the equilibrium Player 2 applies the following upper boundary strategy*

$$U(\pi) = \frac{K}{\pi} \frac{r}{r - \mu}.$$

The lower trigger strategy of Player 1 of the w type is

$$L = \begin{cases} x_0 & if\ \Gamma_3(x_0) \le 0, \\ \begin{cases} x_0 & with\ prob.\ p \\ 0 & with\ prob.\ (1 - p) \end{cases} & if\ \Gamma_2(x_0) < 0 < \Gamma_3(x_0), \\ 0 & if\ \Gamma_2(x_0) \ge 0, \end{cases}$$

where

$$\Gamma_2(x) = -\frac{c}{r} - \left(\frac{x}{U(\pi_0)}\right)^{\frac{r}{\mu}} \left(1 - \frac{c}{r}\right) + \left(\frac{x}{U(1)}\right)^{\frac{r}{\mu}},$$

$$\Gamma_3(x) = -\frac{c}{r} + \left(\frac{x}{U(1)}\right)^{\frac{r}{\mu}},$$

and

$$p = \frac{1 - N\pi_0}{(1 - N)\pi_0},$$

with

$$N = \frac{r - \mu}{r} \frac{x_0}{K} \left[\frac{r - c}{r \left(\frac{x_0}{U(1)}\right)^{\frac{r}{\mu}} - c}\right]^{\frac{\mu}{r}}.$$

$\Gamma_2(x)$ and $\Gamma_3(x)$ characterize the incentive constraints of Player 1. $\Gamma_2(x_0)$ is the difference between the total cost of signaling with entry at $x = U(\pi_0)$ and revealing immediately with entry at $x = U(1)$. $\Gamma_3(x_0)$ is the difference between the cost of continuous signaling without entry and revealing immediately.

The reason that the w type randomizes at $t = 0$ for some intermediate parameter values is intuitively clear. Because x increases over time, Player 2 ultimately contests the stake under a pooling outcome. Taking this into account Player 1 does not choose (a pure strategy) signaling if $\Gamma_2(x_0) < 0$. Under the full separating outcome Player 2 never enters if he expects to face the s type. Then, however, the w type may be tempted to deviate and imitate the s type if $\Gamma_3(x_0) > 0$. As $\Gamma_2(x_0) < \Gamma_3(x_0)$, there are parameter values at which neither pure pooling nor pure separating is an equilibrium outcome. The probabilities p and $(1 - p)$ follow from Bayes rule and the indifference of Player 1 for signaling and revealing at $t = 0$.

Finally, we look at the simplest case when X_t is constant over time, i.e. $\mu = \sigma = 0$. The signaling game simplifies to a repeated game under stationary conditions. It is not difficult to derive the following result as the middle ground of the two cases described in Propositions 2.1 and 2.2.

Corollary 2.3 *If $\mu = \sigma = 0$, the unique MPBE is*
(i) if $x_0 \leq K$, Player 1 of the w type immediately reveals and Player 2 never enters;
(ii) if $\pi_0 x_0 \leq K < x_0$ and $c \leq r$, Player 1 never reveals and Player 2 never enters;
(iii) in the remaining case, Player 1 does not signal and Player 2 enters immediately against the w type and never enters against the s type.

2.4 Equilibrium analysis

2.4.1 Preliminaries

We begin with pointing out two key implications of the stochastic state variable to the signaling game. Let us denote by $\Gamma(x)$ the difference between the payoffs of the w type from continuous signaling (up to entry at $U(\pi)$ when Player 1 incurs a loss of -1) and revealing immediately (that triggers entry at $U(1)$), that is

$$\Gamma(x) = -\frac{c}{r} - \left(\frac{x}{U(\pi)}\right)^{\beta_1}\left(1 - \frac{c}{r}\right) + \left(\frac{x}{U(1)}\right)^{\beta_1}.$$

For any $x \leq U(\pi)$, $\Gamma(x)$ captures the incentive compatibility constraint of the w type for signaling.[3] Analyzing the expression we observe that as long as $U(\pi) \geq U(1)$ then $\Gamma'(x) > 0$ for all $x \geq 0$, and $\Gamma(y) = 0$ for some $y > 0$ (the condition $U(\pi) \geq U(1)$ intuitively holds, and we shall see later it is always true in equilibrium). The implication is that if Player's 1 incentive constraint for signaling is satisfied at $t = 0$, then it will be binding whenever x exceeds x_0. However if at some time x falls below x_0, then Player 1 of the w type might prefer to stop signaling, reveal her type and wait for Player 2 to take his prize at $U(1)$. The fact that in the case $\sigma = 0$ and $\mu \geq 0$ the incentive constraint remains to be satisfied if it is satisfied at x_0, made the equilibrium strategies in the deterministic case relatively simple. Now, however, in the general case, the weak type of Player 1 decides when to stop signaling at lower trigger L taking into account its strategic effect on the entry decision of Player 2. Similarly, Player 2's choice of U is strategic with respect to Player 1's revelation decision.

The second key feature of the stochastic case is that, except for some extreme starting values, the signaling game has no MPBE in pure strategies.

---

[3] For simplicy of exposition we only consider the case that for a given (x, π), $x \leq U(\pi)$. More generally, $U(\pi)$ could be replaced by $\bar{U}(\pi) = \max\{x, U(\pi)\}$ without bringing new insights but complicating notation.

Lemma 2.4 (No pure strategy equilibrium) *If the game is not stopped at $t = 0$, then there is no MPBE in pure strategies.*

The lemma is a special case of Lemma 2.8 below, hence we postpone the formal proof to that point. The intuition for the result is as follows. Player 2 seeing no action at the supposed equilibrium pure strategy L, updates his belief to 0 and never contests ($U(0) = \infty$). But then Player 1 of the weak type deviating from the equilibrium (not revealing at L) obtains a discrete increase in the value function, which upsets the proposed equilibrium. In other words, if L is a best response to $U(\pi)$ for a given (x, π), $\pi > 0$, it is no longer a best response to $U(0)$ at $(x, 0)$.

A similar intuition that explains the nonexistence of a pure strategy equilibrium leads to the anticipation that the distribution G for the mixed strategy of Player 1 should have no atoms. In the remainder of this subsection we prove in a number of steps that there are no gaps in the distribution G and no pure strategies are chosen with positive probability, except possibly at x_0.

Let us denote the support of a distribution G by $\text{supp}(G)$ (that is, the smallest closed set such that the distribution G assigns zero probability to all events not in this set). The question we ask first is what $\tilde{U}(l)$ must be, so that Player 1 chooses l in the support of G. We use the requirement that if $l \in \text{supp}(G)$, then Player 1 must be indifferent to revealing when the minimum l is reached for the first time.

Let $F(x, m)$ denote the expected discounted value of Player 1 of the w type such that Player 1 is indifferent between stopping and continuing at $x = m$ for all $m \in \text{supp}(G)$. In the continuation region, for $x \in (m, \tilde{U}(m))$ with m fixed, the following Bellman-type equation holds:

$$rF(x, m) = \mu x F_x(x, m) + \frac{1}{2}\sigma^2 x^2 F_{xx}(x, m) - c. \tag{2.8}$$

Note that m, the second dimension of the state space, does not appear directly in the differential equation. The reason is that m does not change during an infinitesimal time interval if $x > m$. The general solution to the differential equation is

$$F(x, m) = B_1(m)x^{\beta_1} + B_2(m)x^{\beta_2} - \frac{c}{r}, \tag{2.9}$$

where β_1 and β_2 are the roots of characteristic quadratic $\frac{1}{2}\sigma^2\beta(\beta - 1) + \mu\beta - r = 0$, positive and negative, respectively. The continuous and smooth fit principles at the boundaries give the following conditions:

$$F(\tilde{U}(m), m) = -1, \tag{2.10}$$

$$F(m, m) = \Omega(m), \tag{2.11}$$

$$F_x(m, m) = \Omega'(m), \tag{2.12}$$

for all $m \in \text{supp}(G)$. Condition (2.10) states that the continuation value equals the terminal payoff -1 at $\tilde{U}(m)$. $\Omega(m)$ is the payoff of the w type if the type is revealed at m and its value is given in (2.7). Conditions (2.11)-(2.12) reflect that Player 1 is indifferent between revealing and not revealing. Additionally the normal reflection condition[4] holds at $x = m$:

$$F_m(m, m) = 0. \tag{2.13}$$

Let us define $L_1 = \inf(\text{supp}(G))$ to be the infimum of the set $\text{supp}(G)$. The forgoing arguments can be used to characterize L_1. This is the threshold at which Player 1 optimally reveals with probability one under the most favorable circumstances, i.e. when Player 2 believes that $\theta = s$ and never enters. The differential equation (2.8) is then coupled with the value matching and smooth pasting conditions $F(L_1, L_1) = \Omega(L_1)$ and $F_x(L_1, L_1) = \Omega'(L_1)$. Later we shall also discuss the supremum L_0 of the set $\text{supp}(G)$, that is $L_0 = \sup(\text{supp}(G))$. Furthermore, let $\tilde{U}_G : (L_1, U(1)) \to \mathbb{R}_+$ be the solution to (2.14) below. The next lemma provides the condition for m to be in $\text{supp}(G)$, characterizes L_1 and some properties of $\tilde{U}_G$. To shorten notation let $\delta_1(m) = (\tilde{U}(m)/m)^{\beta_1}$ and $\delta_2(m) = (\tilde{U}(m)/m)^{\beta_2}$.

Lemma 2.5 (Support of G) *(i) If $m \in \text{supp}(G)$, then $\tilde{U}(m)$ satisfies*

$$\left[(\beta_1 - \beta_2)\,\Omega(m) - \frac{c}{r}\beta_2\right]\delta_1(m) + \frac{c}{r}\beta_1\delta_2(m) - \left(\frac{c}{r} - 1\right)(\beta_1 - \beta_2) = 0. \tag{2.14}$$

(ii) It holds that

$$L_1 = \left(\frac{\beta_2}{\beta_2 - \beta_1}\frac{c}{r}\right)^{\frac{1}{\beta_1}}\frac{K\beta_1}{\beta_1 - 1}.$$

(iii) For any $m \in (L_1, U(1))$ let $\tilde{U}_G(m)$ denote the unique positive solution of (2.14). $\tilde{U}_G : (L_1, U(1)) \to \mathbb{R}_+$ is continuous and strictly decreasing.

Proof. (i) Combining conditions (2.11) and (2.12) with the general solution (2.9) we obtain

$$B_1(m) = \frac{1}{\beta_1 - \beta_2}\left[(\beta_1 - \beta_2)\,\Omega(m) - \frac{c}{r}\beta_2\right]m^{-\beta_1},$$

$$B_2(m) = \frac{c}{r}\frac{\beta_1}{\beta_1 - \beta_2}m^{-\beta_2},$$

---

[4] The *normal reflection* conditions are used in the optimal stopping problems involving a extemum (maximum or minimum) process. For a formal verification that (2.8) together with a boundary condition corresponding to (2.13) hold in problems involving the minimum, see Peskir (1998) (see also Peskir and Shiryaev (2006, Ch. 13)).

for all $m \in \text{supp}(G)$. We note that after substituting $B_1(m)$ and $B_2(m)$ into (2.13), (2.13) holds as an identity. Then substituting $B_1(m)$ and $B_2(m)$ in (2.10) yields (2.14), where the dependence of the functions on m is omitted for brevity.

(ii) In addition to $F(L_1, L_1) = \Omega(L_1)$ and $F_x(L_1, L_1) = \Omega'(L_1)$, the value must be bounded as x goes to infinity. Applying all three conditions to the general solution of (2.8) yields part (ii).

(iii) Clearly, it must hold that $\tilde{U}(m) \geq m$. Then the derivative of the left hand side of (2.14) with respect to $\tilde{U}$ is

$$
\begin{aligned}
\left[\beta_1 (\beta_1 - \beta_2) \delta_1(m)\Omega(m) - \frac{c}{r}\beta_1\beta_2 (\delta_1(m) - \delta_2(m)) \right] \tilde{U}(m)^{-1} &< \frac{c}{r}\beta_1\beta_2\delta_2(m)\tilde{U}(m)^{-1} \\
&< 0,
\end{aligned}
$$

where the inequality follows from the fact that $\Omega(m) < -\frac{\beta_2}{\beta_2 - \beta_1}\frac{c}{r}$ for $m > L_1$ (using part (ii) of the lemma). Moreover, observing that the left hand side of (2.14) is continuous in positive $\tilde{U}$ and it diverges to $+\infty$ at $\tilde{U} = 0$ and diverges to $-\infty$ as $\tilde{U}$ goes to infinity, we conclude that there is a unique positive root $\tilde{U}_G$. As the left hand side of (2.14) is strictly positive for $\tilde{U}(m) = m$ (with $m \in (L_1, U(1))$), it follows that $\tilde{U}_G(m) > m$, as expected. Finally, a straightforward application of the implicit function theorem and some algebra delivers that $\tilde{U}_G(\cdot)$ is strictly decreasing. ∎

Equation (2.14) characterizes the entry boundary function $\tilde{U}_G(m)$ of Player 2 that would make m a part of the mixed strategy of Player 1. In other words, Lemma 2.5 means that

$$
\begin{aligned}
F(m, m) < \Omega(m) \quad \text{if } \tilde{U}(m) < \tilde{U}_G(m), \\
F(m, m) > \Omega(m) \quad \text{if } \tilde{U}(m) > \tilde{U}_G(m).
\end{aligned}
\tag{2.15}
$$

If $\tilde{U}(m) < \tilde{U}_G(m)$ then Player 1 strictly prefers revealing at m than continuing. If $\tilde{U}(m) > \tilde{U}_G(m)$ then Player 1 prefers continuing at m. The relevant domain of $\tilde{U}_G$ is $(L_1, U(1))$. At L_1, $\tilde{U}_G$ diverges to infinity (by construction of L_1). No minima above $U(1)$ can be in the support of G and $\tilde{U}_G$ approaches $U(1)$ at $m = U(1)$. The intuition is as follows. If Player 1 reveals at $U(1)$ she faces an immediate contestant. As we assume that $c < r$, Player 1 always prefers postponing immediate entry, so she is indifferent to revelation only if continuation also leads to immediate entry, i.e. if $\tilde{U}_G(m) = U(1)$.

It is important to realize that L_1 can be characterized based solely on the problem of Player 1 as the choice of L_1 is a nonstrategic decision. The equilibrium distribution G and in particular the level of L_0 must incorporate the strategic effects on the behavior of Player 2.

Let us denote the closure of the set where $\tilde{U}$ is not constant by $S(\tilde{U})$.

Lemma 2.6 (Common support) $\text{supp}(G) = S(\tilde{U})$.

Proof. Suppose that $x \in \text{supp}(G)$ and $x \notin S(\tilde{U})$. It means that there is an open neighborhood of x, $b(x)$, such that for all $x' \in b(x)$, $\tilde{U}(x') = \tilde{U}(x)$. But then at any $x' - \epsilon < x$, with some $\epsilon > 0$ and $x' \in b(x)$, Player 1, by Lemma 2.5, strictly prefers revealing than continuing. So Player 1 reveals with probability 1 at $x' - \epsilon$ and then $\tilde{U}(x') = \infty \neq \tilde{U}(x)$. Consequently it can not be that $x \in \text{supp}(G)$ and $x \notin S(\tilde{U})$.

Next suppose that $x \notin \text{supp}(G)$ and $x \in S(\tilde{U})$. It means that there is an open neighborhood of x, $b(x)$, such that for all $x' \in b(x)$, $G(x') = G(x)$. This means also that $x' \in b(x)$ the belief π cannot be different as there is no information revealed. As for all $x' \in b(x)$ the belief is the same and the probability of reaching minima in the support of G are the same, it can not be optimal to play $\tilde{U}(x') \neq \tilde{U}(x)$. Thus $b(x) \nsubseteq S(\tilde{U})$ contradicting that $x \in S(\tilde{U})$. ∎

Lemma 2.6 means that $\tilde{U}$ is not constant only on the set that is in the support of G. This observation is used to prove the next lemma.

Lemma 2.7 (No gaps) *There are no gaps in* $\text{supp}(G)$.

Proof. Suppose that there is a gap (a, b) over which G is constant and a and b belong to the support of G (recall that $\text{supp}(G)$ is a closed set). Then, by Lemma 2.5, $\tilde{U}(a) = \tilde{U}_G(a)$ and $\tilde{U}(b) = \tilde{U}_G(b)$. By Lemma 2.6, $\tilde{U}$ is constant over (a, b). If Player 1 does not put any positive probability on strategies in (a, b), $\tilde{U}(l)$ must be larger than or equal to $\tilde{U}_G(l)$ by (2.15). Then, as $\tilde{U}$ is constant in (a, b) and $\tilde{U}_G$ is strictly decreasing (Lemma 2.5(iii)), it follows that $\tilde{U}(l) = \tilde{U}(a) > \tilde{U}_G(l)$ for all $l \in (a, b)$. Thus, by (2.15), the continuation payoff $F(l, l)$ is strictly larger than terminal payoff $\Omega(l)$ for all $l \in (a, b)$. We also note that $\Omega(\cdot)$ is a continuous function. Then the following inequality holds

$$F(a, a) = \Omega(a) = \lim_{a' \uparrow a} \Omega(a') < \lim_{a' \uparrow a} F(a', a').$$

But then by an infinitesimal deviation Player 1 gets a benefit that is bounded away from zero, so a cannot be in G. Consequently there cannot be gaps in $\text{supp}(G)$. ∎

Finally we show that G has no atoms. To prove it let us denote the probability that Player 1 stops exactly at stopping time corresponding to trigger strategy x by $J(x)$.

Lemma 2.8 (No atoms) *For* $x \in [L_1, x_0)$, $J(x) = 0$.

Proof. Suppose $J(l) = p > 0$ for some $l \in (0, x_0)$. As $l \in \text{supp}(G)$, $F(l, l) = \Omega(l)$ and $\tilde{U}(l) = \tilde{U}_G(l)$ by Lemma 2.5. Suppose that at the time l is reached for the first time, the belief is $\pi \in (0, 1)$. If Player 1 does not stop, Player 2 uses Bayes rule to update his belief to $\pi' = \frac{(1-p)\pi}{1 - \pi p} < \pi$. Let us denote $\lim_{l' \uparrow l} \tilde{U}(l')$ by $\tilde{U}^-(l)$ and $\lim_{l' \uparrow l} \tilde{U}_G(l')$ by $\tilde{U}_G^-(l)$. We have two cases to consider.

Case 1. Suppose that $U(\pi') > U(\pi)$, that is $\tilde{U}^-(l) > \tilde{U}(l)$. By Lemma 2.5(iii), U_G is continuous, so $\tilde{U}^-(l) > \tilde{U}_G^-(l)$. But then by the same argument as in Lemma 2.7,

Player 1 faces a jump in the value and by an infinitesimal deviation gets a benefit that is bounded away from zero. Formally, the following inequality holds:

$$F(l, l) = \Omega(l) = \lim_{l' \uparrow l} \Omega(l') < \lim_{l' \uparrow l} F(l', l').$$

Hence l is not played in the mixed strategy G and $J(l) = 0$.

   Case 2. Suppose that $U(\pi') \leq U(\pi)$, that is $\tilde{U}^-(l) \leq \tilde{U}(l)$. We shall show that Player 2's best response is never $U(\pi') \leq U(\pi)$ if $p > 0$. To do this let us consider Player 2's best response problem given G and $J(l) = p > 0$. Let $V(x, l)$ be the value of Player 2 in this best response problem. In the continuation region, that is for $x \in (l, \tilde{U}(l))$, $V(x, l)$ must satisfy the Bellman-type equation

$$rV(x, l) = \mu x V_x(x, l) + \frac{1}{2}\sigma^2 x^2 V_{xx}(x, l).$$

At the boundaries the following conditions hold

$$
\begin{aligned}
V(\tilde{U}(l), l) &= \pi \tilde{U}(l) - K, \\
V_x(\tilde{U}(l), l) &= \pi, \\
V(l, l) &= \pi p W(l) + (1 - \pi p) V^-(l, l),
\end{aligned}
\tag{2.16}
$$

where $V^-(l, l) = \lim_{l' \uparrow l} V(l', l')$ denotes the continuation value just after the minimum at l is reached. Let $\Delta_1 = (l/\tilde{U}(l))^{\beta_1} - (l/\tilde{U}(l))^{\beta_2}$ and $\Delta_2 = \beta_2(l/\tilde{U}(l))^{\beta_1} - \beta_1(l/\tilde{U}(l))^{\beta_2}$. Then using the general solution to the differential equation and the two first boundary conditions we obtain that

$$V(l, l) = \frac{1}{\beta_1 - \beta_2}\left[\pi\tilde{U}(l)(\Delta_1 - \Delta_2) + K\Delta_2\right]. \tag{2.17}$$

$V^-(x, l)$ must satisfy the same differential equation, but the boundary conditions at $\tilde{U}^-(l)$ become

$$
\begin{aligned}
V^-(\tilde{U}^-(l), l) &= \pi'\tilde{U}^-(l) - K, \\
V_x^-(\tilde{U}^-(l), l) &= \pi'.
\end{aligned}
$$

Solving for $V^-(x, l)$ we find

$$V^-(l, l) = \frac{1}{\beta_1 - \beta_2}\left[\pi'\tilde{U}^-(l)(\Delta_1^- - \Delta_2^-) + K\Delta_2^-\right],$$

where $\Delta_1^- = (l/\tilde{U}^-(l))^{\beta_1} - (l/\tilde{U}^-(l))^{\beta_2}$ and $\Delta_2^- = \beta_2(l/\tilde{U}^-(l))^{\beta_1} - \beta_1(l/\tilde{U}^-(l))^{\beta_2}$. As $V^-(l,l)$ increases in $\tilde{U}$ (as long as $\pi'\tilde{U}^-(l) \geq \frac{\beta_2}{\beta_2-1}U(1)$,but this must be the case in the best response of Player 2, as the point with equality is where $V^-(l,l)$), we can write the following inequality

$$V^-(l,l) \leq \frac{1}{\beta_1 - \beta_2} \left[\pi'\tilde{U}(l)(\Delta_1^- - \Delta_2^-) + K\Delta_2^- \right].$$

Substituting this inequality together with (2.17) and $\pi' = \frac{(1-p)\pi}{1-\pi p}$ in (2.16) yields

$$p\frac{1}{\beta_1 - \beta_2} \left[\tilde{U}(l)(\Delta_1 - \Delta_2) + K\Delta_2 \right] \leq pW(l). \tag{2.18}$$

The term on the left hand side equals $pW(l)$ if $\tilde{U}(l) = U(1)$ and increases in $\tilde{U}(l)$ if $\tilde{U}(l) > U(1)$. As $l \in \mathrm{supp}\,(G)$ then, by Lemma 2.5, $\tilde{U}(l) > U(1)$, hence the weak inequality (2.18) holds only if $p = 0$.

Note that the argument in Case 1 does not hold for the initial state (x_0, x_0) at $t = 0$. At (x_0, x_0) it can be that $F(x_0, x_0) < \Omega(x_0)$. In this case Player 1 prefers revealing above continuing and reveals with probability 1 if $x_0 \leq L_1$ or randomizes between continuing and revealing if $x_0 > L_1$. ∎

2.4.2   Learning and best response of the uninformed player

In this section we analyze the optimal strategy U of Player 2 given that Player 1 adopts an arbitrary continuous strategy G. As Player 1 randomizes over the lower trigger strategies, Player 2, while observing that the game is not stopped at a new minimum in the support of G, updates its belief about θ. Thus, provided that the signaling game is not stopped by time t, the belief π_t will depend on the running minimum M_t of X_t. The learning process is described by a function $\Pi(m)$ derived by Bayes rule

$$\Pi(m) = \frac{(1 - G(m))\,\pi_0}{1 - G(m)\pi_0},$$

such that $\pi_t = \Pi(M_t)$. In a similar fashion, the minimum process is used to update the distribution of the mixed strategy of the w type. Given a minimum m let $g(x; m) = G'(x)/(1 - G(m))$ be an updated density of G at $x \leq m$ conditional on $\theta = w$. For brevity we denote $g(m; m)$, the hazard function of G at m, by $g(m)$.

Player 2 chooses a stopping boundary $\tilde{U}$ in the state space $\{(x, m) \in \mathbb{R}_+^2 : x \geq m\}$ to maximize his expected discounted value taking into account, firstly, the possibility of learning the w type if the type is revealed at a random lower trigger and, secondly, the gradual learning about θ if new values in the support of G are reached. Denote

this value by $V(x,m)$. The dependence on the running minimum places the problem of Player 2 in line with some work on (non-standard) lookback options (e.g. Guo and Shepp (2001)) and more general recent literature on optimal stopping of the maximum process (Peskir (1998)). Using the dynamic programming arguments and Itô's lemma, we derive that in the continuation region, for $x \in (m, \tilde{U}(m))$ with fixed m, $V(x,m)$ must satisfy the ordinary differential equation

$$rV(x,m) = \mu x V_x(x,m) + \frac{1}{2}\sigma^2 x^2 V_{xx}(x,m). \tag{2.19}$$

Note that, similar to differential equation (2.8) in the problem of the informed player, derivatives in m do not appear in equation (2.19). In the space $\{(x,m) \in \mathbb{R}_+^2 : x \geq m\}$, m changes only after hitting the diagonal $x = m$, and this property shall be employed below in the boundary condition (2.22).

The general solution to (2.19) is of the form

$$V(x,m) = A_1(m)x^{\beta_1} + A_2(m)x^{\beta_2}.$$

The coefficients $A_1(m)$ and $A_2(m)$ as well as the optimal boundary $\tilde{U}(m)$ are determined by considering extremes in the continuation region in (x,m). At the boundary $x = \tilde{U}(m)$ between the continuation and stopping region we require the familiar conditions of continuous and smooth fit, that is

$$V(\tilde{U}(m), m) = \Pi(m)\tilde{U}(m) - K, \tag{2.20}$$

$$V_x(\tilde{U}(m), m) = \Pi(m). \tag{2.21}$$

When $x = m$, that is on the diagonal in $\mathbb{R}_+^2$, the probability of facing the w type is $\Pi(m)$ and, upon a marginal change in m, the probability that the w type reveals is $-g(m)$. In Section 2.3.2 we derived that, if the w type reveals, Player 2 gets $W(m)$ given by (2.6). It follows that at $x = m$ it holds that[5]

$$V_m(m,m) = \Pi(m)g(m)\left(V(m,m) - W(m)\right). \tag{2.22}$$

To shorten notation let $\Delta_1 = \Delta_1(m, \tilde{U}(m)) = (m/\tilde{U}(m))^{\beta_1} - (m/\tilde{U}(m))^{\beta_2}$ and $\Delta_2 = \Delta_2(m, \tilde{U}(m)) = \beta_2(m/\tilde{U}(m))^{\beta_1} - \beta_1(m/\tilde{U}(m))^{\beta_2}$. Then after solving (2.19)-(2.22) we obtain that the best response strategy of Player 2 is given by the following

---

[5] The boundary condition (2.22) closely corresponds with the normal reflection conditions in the standard optimal stopping problems of maximum (or minimum) process (see (2.13) and footnote 4); it reduces to $V_m(m,m) = 0$ if no event happens at m (that is, if either $\Pi(m)g(m) = 0$ or $V(m,m) = W(m)$).

differential equation (after suppressing the dependence of the functions on m)

$$\tilde{U}'[(1 - \beta_1)(1 - \beta_2)\Pi\tilde{U} - \beta_1\beta_2 K]\Delta_1$$
$$= -\Pi'\tilde{U}^2(\Delta_1 - \Delta_2) - g\Pi\tilde{U}[(\beta_1 - \beta_2)W - \Pi\tilde{U}(\Delta_1 - \Delta_2) - K\Delta_2]. \quad (2.23)$$

The term on the left-hand side stems from the dependence of the terminal payoff at the upper boundary on the minimum process and is standard in problems of optimal stopping of extremum processes (cf. equation (6) in Guo and Shepp (2001)). The first term on the right-hand side comes from the effect of learning from the minimum process. The second term on the right-hand side captures the influence of the type revelation at a random lower trigger.

To identify a relevant boundary condition, we observe that when m reaches L_1, which is the lower bound on the support of G, the belief that $\theta = w$ is zero. At this point Player 2 never enters and thus the boundary condition at $m = L_1$ is

$$\tilde{U}(L_1) = U(0) = \infty.$$

2.4.3   Equilibrium

The analysis of both players' strategies of the previous sections provides ingredients for the equilibrium result stated in the proposition below. For technical reasons we shall formulate and prove the proposition for a 'perturbed' version of the game in which Player 2 gets εX_t if he contests the stake at time t from the strong type. In other words, the strong player is not a 'natural monopolist'. The fraction $\varepsilon > 0$ is assumed to be small (in the baseline model $\varepsilon = 0$). Our choice of the case $\varepsilon = 0$ so far stems from the attempt to simplify the exposition. On the other hand, this section shows that a small perturbation $\varepsilon > 0$ readily delivers the uniqueness of the equilibrium strategies.[6]

Assume therefore that $\varepsilon > 0$. The analysis of the previous sections can be accordingly adjusted without much difficulty. In particular, the lower bound on the support of G depends on ε and we denote it L_1^ε with $L_1^0 = L_1$. Then after denoting the strategies of Player 2 in the perturbed model by $\tilde{U}^\varepsilon$ and U^ε, we obtain $\tilde{U}^\varepsilon(L_1^\varepsilon) = U^\varepsilon(0) = \frac{1}{\varepsilon}K\beta_1/(\beta_1 - 1)$ (the derivation is similar to the one in Section 2.3.1). The boundedness of $U^\varepsilon(0)$ if $\varepsilon > 0$ is the effect of the ε-perturbation ensuring the uniqueness in the statement of the proposition. For simplicity, we suppress superscript ε from now on.

---

[6]Similar types of assumptions to ensure equilibrium uniqueness have been used in other contexts. For example, in a war of attrition with incomplete information, Fudenberg and Tirole (1986) characterize the equilibrium strategies, as we do in our model, in terms of differential equations. The solution to the differential equation and thus the equilibrium are unique if the firms are not natural monopolists.

The proposition does not include the trivial situation if $x_0 < L_1$, in which case the w type reveals with probability one at $t = 0$.

Proposition 2.9 *Let* $\hat{\Pi} = \Pi + (1 - \Pi)\varepsilon$, $\varepsilon > 0$, *and*

$$f_1(m, G, \tilde{U}) =$$

$$\frac{f_2(1 - \pi_0 G)^2[(1 - \beta_1)(1 - \beta_2)\hat{\Pi}\tilde{U} - \beta_1\beta_2 K]\Delta_1}{(1 - \varepsilon)\pi_0(1 - \pi_0)\tilde{U}^2(\Delta_1 - \Delta_2) - \pi_0(1 - \pi_0 G)\tilde{U}[(\beta_1 - \beta_2)W - \hat{\Pi}\tilde{U}(\Delta_1 - \Delta_2) - K\Delta_2]},$$

$$f_2(m, G, \tilde{U}) = \frac{\beta_2 c\tilde{U}\,(\delta_1 - \delta_2)}{(\beta_1 - \beta_2)\,(c\delta_2 + r - c)\,m},$$

and denote by $(G^*, \tilde{U}^*)$ *the (unique) solution to the system of differential equations*

$$G'(m) = f_1(m, G, \tilde{U}), \ G(L_1) = 1, \tag{2.24}$$

$$\tilde{U}'(m) = f_2(m, G, \tilde{U}), \ \tilde{U}(L_1) = U(0) = \frac{\beta_1}{\beta_1 - 1}\frac{K}{\varepsilon}. \tag{2.25}$$

Then a pair of strategies $(G, \tilde{U})$ *is the unique MPBE of the signaling game if*

$$G(m) = \begin{cases} 1 & if \ m < L_1, \\ G^*(m) & if \ L_1 \leq m \leq L_0, \\ 0 & if \ L_0 < m \leq x_0, \ L_0 \neq x_0. \end{cases}$$

If $L_0 = x_0$, *then at* $t = 0$ *Player 1 of the w type reveals its type with probability* $G^*(L_0)$, *signals with probability* $1 - G^*(L_0)$ *and at* $t > 0$ *plays according to* G. *Player 2 contests at*

$$\tilde{U}(m) = \begin{cases} \tilde{U}^*(L_1) & if \ m < L_1, \\ \tilde{U}^*(m) & if \ L_1 \leq m \leq L_0, \\ \tilde{U}^*(L_0) & if \ L_0 < m \leq x_0. \end{cases}$$

The lower bound on the support of G *is given by the solution* L_1 *of*

$$\left[(\beta_1 - \beta_2)\,\Omega(L_1) - \beta_2\frac{c}{r}\right]\left(\frac{U(0)}{L_1}\right)^{\beta_1} + \beta_1\frac{c}{r}\left(\frac{U(0)}{L_1}\right)^{\beta_2} - (\beta_1 - \beta_2)\left(\frac{c}{r} - 1\right) = 0 \tag{2.26}$$

and the upper bound is

$$L_0 = \min\{\bar{L}_0, x_0\}, \ \bar{L}_0 = \inf\{m \geq L_1 : G^*(m) = 0\}.$$

Finally, $U(1) = K\beta_1/(\beta_1 - 1)$.

Proof. In the proof we refer to the results in Sections 2.3 and 2.4.1-2.4.2 stated there for the limit case $\varepsilon = 0$, while pointing to the necessary adjustments as $\varepsilon > 0$. If $x_0 > L_1$, then by Lemma 2.4 there exist no equilibrium in pure strategies. By Lemma 2.8, the mixed strategy G of Player 1 has to be a continuous function with a support on some $[L_1, L_0]$. L_1 is given in Lemma 2.5(ii) in the limit case $\varepsilon = 0$. If $\varepsilon > 0$, a similar derivation, but with an additional boundary condition

$$F(U(0), L_1) = -1,$$

yields the implicit equation for L_1 as stated in the proposition. From Lemma 2.5, $m \in (L_1, x_0)$ is in the support of G only if $\tilde{U}(m)$ satisfies (2.14), which after differentiation gives $\tilde{U}'(m) = f_2(m, G, \tilde{U})$, with initial value condition $\tilde{U}(L_1) = \frac{1}{\varepsilon} K \beta_1 / (\beta_1 - 1)$.

From Section 2.4.2 it follows that Player 2 chooses the given $\tilde{U}(m)$ if $G(m)$ satisfies equation (2.23). If $\varepsilon > 0$, the boundary conditions (2.20) and (2.21) are substituted with

$$
\begin{aligned}
V(\tilde{U}(m), m) &= \hat{\Pi}(m)\tilde{U}(m) - K, \\
V_x(\tilde{U}(m), m) &= \hat{\Pi}(m).
\end{aligned}
$$

Combination of these and (2.22) with the solution to (2.19) and some reorganization yield $G'(m) = f_1(m, G, \tilde{U})$. The initial value condition $G(L_1) = 1$ follows from the construction of L_1.

The upper bound on the support of G is then $L_0 = \min\{\bar{L}_0, x_0\}$, with $\bar{L}_0 = \inf \{m \geq L_1 : G^*(m) = 0\}$. If $L_0 < x_0$, then clearly $G(m) = 0$ and $\tilde{U}(m) = \tilde{U}(L_0)$ for $m > L_0$. If $\bar{L}_0 < x_0$, then neither continuing nor stopping is a pure strategy equilibrium at $t = 0$ (with the arguments parallel to those in Lemma 2.4). In the mixed strategy at $t = 0$, Player 1 randomizes to be indifferent between revealing and signaling and chooses probabilities $G^*(L_0)$ and $1 - G^*(L_0)$, respectively.

The complete information threshold $U(1)$ is derived in Section 2.3.1.

Finally, we confirm the uniqueness of the equilibrium. It is not difficult to verify that the initial value problem (2.24)-(2.25) satisfies the Lipschitz condition as long as $\varepsilon > 0$. Thus there is a unique solution $(G^*, \tilde{U}^*)$ if the informed player applies a continuous strategy G (except at $t = 0$). As we have shown in Section 2.4.1 this is the only kind of strategy played in equilibrium. ∎

A graphical representation of the equilibrium strategies is shown in Figure 2. The result can be interpreted as follows. For the outcome to satisfy the subgame perfectness criterion, Player 1 reveals at a random lower trigger with a continuous distribution G on some $[L_1, L_0]$. Player 2, while observing new minima reached in the support of G
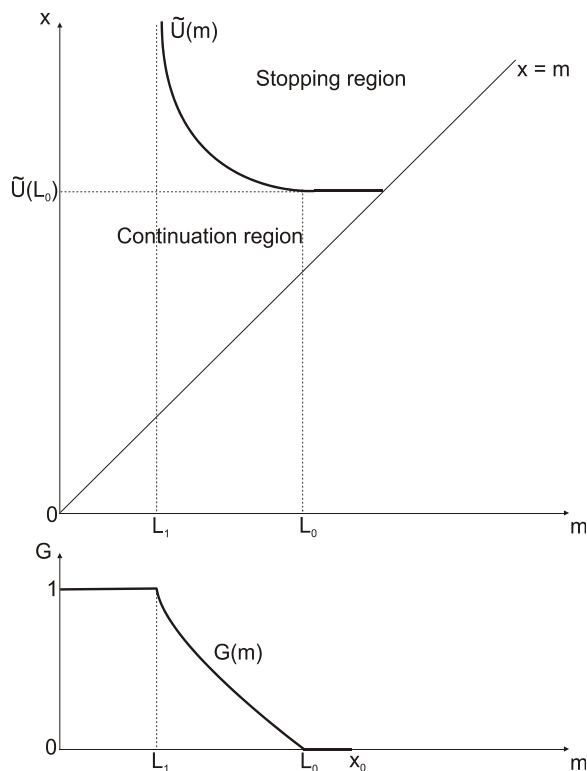
FIGURE 2.2. Equilibrium strategies.

without the type revealed, updates his belief that he is facing the weak type (belief π decreases). Consequently, with decreasing m Player 2 requires a higher x to risk contesting the stake so $\tilde{U}(m)$ rises. Given that the game has not been stopped by the time x reaches L_1 (with $G(L_1) = 1$), the true type must be s and Player 2 contests at $U(0)$. The equilibrium slope of G in (2.24) is such that Player 2's best response to this G is $\tilde{U}$ (given by (2.25)) such that it is indeed optimal for Player 1 to choose a continuous G with support on $[L_1, L_0]$.

We would like to comment on the equilibrium in the limit as ε goes to zero. The problem is that the initial point $m = L_1$ becomes a singular point for the differential system (the system does satisfy the Lipschitz continuity condition) and thus it has no unique solution. Such indeterminacy is not unusual in optimal stopping problems with extremum process. In a related setup with similar indeterminacy, Peskir (1998) introduced the maximality principle that determines the optimal trajectory of the optimal stopping boundary. Developing a corresponding principle in our strategic problem is beyond the scope of this chapter. Nevertheless, building on the analogy to the maximality principle, we may anticipate that in our setup the equilibrium distribution $G^*(m)$ would be determined by choosing the maximal one among those solutions to (2.24) that are non-decreasing.

Our analysis of the two-type case can be adapted to the case of a continuum of types of the informed player. Such a setup would ease one technical issue, namely the informed player would not necessarily need to apply mixed strategies. On the other hand, with a continuum of types it might be more demanding to sustain the initial pooling. While in a binary signaling space, initial pooling strategy may be incentive compatible even with a continuum of types, typically we would not expect it in the case of a continuous signaling space (say in $\mathbb{R}$). The latter is the case in the application of the model to limit pricing in the next section, hence our focus on the two-type case, but in general the modeling will depend on the application in mind.

2.5   Example: Limit pricing

In this section we apply the dynamic stochastic signaling setup to a standard signaling situation known from industrial organization, namely limit pricing under incomplete information. Following Milgrom and Roberts (1982) we assume that the incumbent's cost is not directly observable by the potential entrant. When threatened by entry, the weak incumbent may, by setting low prices, pretend to be a strong one and thus discourage entry. Unlike the existing literature we study the dynamics of signaling and entry in a stochastic market.

The analysis here serves two purposes. First, it demonstrates how the general game of the preceding sections can be adapted to other functional forms of the players' payoffs. Specifically, the cost of signaling and payoffs of the informed player need to be functions of the driving diffusion process (here, the demand shock). Second, the model bridges the gap between the older non-game-theoretic literature on limit pricing, that was often dynamic and considered stochastic markets (Kamien and Schwartz (1971), Gaskins (1971) and Flaherty (1980)), and the game-theoretic equilibrium limit pricing that to a large degree invalidated the older explanations, but left us essentially with one-shot deterministic models.

2.5.1   Setup

We begin by describing the model setup. The incumbent firm, denoted by index 1, already operates in the market. Its profits depend on four factors. Firstly, the profitability of the whole market evolves with a stochastic state variable Y following a geometric Brownian motion. Secondly, denoting the incumbent's cost type by $\theta \in \{w, s\}$, the incumbent's technology may be of low marginal cost C_1^s or high cost C_1^w per unit of time, $C_1^s < C_1^w$. Thirdly, the incumbent may choose other than its monopoly price or quantity to imitate the behavior of another cost type. And lastly, profits depend on

the presence of the entrant firm (firm 2). The entrant's marginal cost C_2 is known with certainty. We assume that upon entry the two firms compete in quantities in Cournot fashion. Below we show that these requirements on the profit flow function can be effectively captured by choosing an appropriate multiplicative constant for a stochastic state variable.

We assume the demand function is isoelastic and is subject to stochastic shocks. Specifically, the inverse demand function of total output Q at time t is given by

$$P_t(Q) = Y_t Q^{-\frac{1}{\gamma}}.$$

$Y = \{Y_t : t \geq 0\}$ is a stochastic state variable following a geometric Brownian motion with drift μ_Y, volatility σ_Y and a standard Brownian motion Z. γ is the demand elasticity and we assume that $\gamma > 1$.

It is straightforward to derive that the profit flow of the unconstrained (that is, not facing a potential entrant) monopolist of type θ at each state Y_t is

$$\frac{Y_t^{\gamma}}{\gamma} \left(\frac{\gamma C_1^{\theta}}{\gamma - 1} \right)^{1-\gamma}. \tag{2.27}$$

Define now a new variable $X_t = f(Y_t) = (Y_t)^{\gamma}$. By Itô's lemma we have

$$dX_t = f' dY_t + \frac{1}{2} f'' dY_t^2 = \mu X_t dt + \sigma X_t dZ_t,$$

where $\mu = \gamma \mu_Y + \frac{1}{2} \gamma (\gamma - 1) \sigma_Y^2$ and $\sigma = \gamma \sigma_Y$ are constants, and f' and f'' denote the first and second order derivatives. Therefore, X is also a geometric Brownian motion adapted to the same filtration.

Note that $\frac{1}{\gamma} \left(\frac{\gamma C_1^{\theta}}{\gamma-1} \right)^{1-\gamma}$ in (2.27) is constant over time, thus with properly chosen parameters the profit flow in (2.27) may be expressed as a constant times a geometric Brownian motion. A similar equivalence can be shown for profit flows under duopolistic competition and in case the monopolist chooses quantities corresponding to optimal quantities of another cost type (i.e. imitates optimal behavior of another type). In particular, the profit flow of the incumbent of type θ is a product of the market state variable X and a constant equal to either of

$$M^{\theta} = \frac{1}{\gamma} \left(\frac{\gamma C_1^{\theta}}{\gamma - 1} \right)^{1-\gamma}, \tag{2.28}$$

$$M^{\theta}(\tilde{C}) = \frac{\gamma \tilde{C} - (\gamma - 1) C_1^{\theta}}{\gamma - 1} \left(\frac{\gamma \tilde{C}}{\gamma - 1} \right)^{-\gamma}, \tag{2.29}$$

$$
D_1^\theta = \begin{cases} \dfrac{\left[\gamma C_2 - (\gamma-1)C_1^\theta\right]^2}{(2\gamma-1)\left(C_1^\theta + C_2\right)}\left[\dfrac{\gamma\left(C_1^\theta + C_2\right)}{2\gamma-1}\right]^{-\gamma} & \text{if } C_1^\theta < \frac{\gamma}{\gamma-1}C_2, \\ 0 & \text{otherwise,} \end{cases}
\tag{2.30}
$$

depending on whether the incumbent is a monopolist choosing its unconstrained monopoly strategy (M^θ), or if the incumbent is a monopolist imitating the monopoly strategy of a firm with marginal cost $\tilde{C}$ $(M^\theta(\tilde{C}))$, or if the incumbent firm operates in a duopoly (D_1^θ). The relationship between the profit constants is not surprising, namely $M^\theta > D_1^\theta \geq 0$, $M^s > M^w$, $M^\theta > M^\theta(\tilde{C})$ and $D_1^s > D_1^w$ for all θ and all $\tilde{C} \neq C^\theta$. From (2.29) it follows that the θ-type incumbent imitating the pricing strategy of a monopolist with cost $\tilde{C}$, has negative profits if $\tilde{C} < \frac{\gamma-1}{\gamma}C_1^\theta$. Equation (2.30) says that the θ-type incumbent is out of the market after entry if the entrant's cost is less than $\frac{\gamma-1}{\gamma}C_1^\theta$.

The incumbent's type is known to the incumbent firm itself but at the initial point of time the potential entrant does not know it. Instead, the prior probability that $\theta = w$ is π_0 and is known to the entrant. Upon entry firm 2 pays the entry cost of K and learns the cost type of the incumbent. After the entrant has entered the market, its profits are affected by the cost level of the incumbent. Given that firm 1 is of the θ type, firms 2's profit flow after entry can be expressed as the product of X and a constant, where the constant is

$$
D_2^\theta = \begin{cases} \dfrac{\left[\gamma C_1^\theta - (\gamma-1)C_2\right]^2}{(2\gamma-1)\left(C_1^\theta + C_2\right)}\left[\dfrac{\gamma\left(C_1^\theta + C_2\right)}{2\gamma-1}\right]^{-\gamma} & \text{if } C_2 < \frac{\gamma}{\gamma-1}C_1^\theta, \\ 0 & \text{otherwise.} \end{cases}
\tag{2.31}
$$

The lower the incumbent's cost the less profitable is entry, that is $D_2^w > D_2^s$. If the entrant knows the incumbent's type to be θ and $C_2 \geq \frac{\gamma}{\gamma-1}C_1^\theta$, then the entrant cannot make positive profits and thus never enters.

We make the following assumptions on the cost structure. (1) There is profitable entry in the market against the w type: $D_2^w > 0$. (2) The incumbent of type w prefers imitating the s-type above facing immediate entry: $M^w(C_1^s) \geq D_1^w$. (3) The incumbent of type s has its marginal cost slightly higher than that of natural monopolist: $C_1^s = (1+\bar{\varepsilon})\frac{\gamma-1}{\gamma}C_2$, where $\bar{\varepsilon}$ is assumed to be a small positive number. Assumption (1) ensures that the game is interesting. Assumption (2) is a necessary condition for the incumbent of type w to engage in signaling (it corresponds the assumption that $c < r$ in the generic game). By assumption (3) we concentrate our attention on the case corresponding to the analysis in Sections 2.2-2.4. The strong incumbent is strong enough to have no incentives to signal its type, yet the entrant gets a small share of the market after entry to ensure uniqueness of the equilibrium strategies (analogously to the ε-perturbation

in Section 2.4.3). Assumption (3) can be rewritten in terms of entrant's profits as $D_2^s = \varepsilon D_2^w$ for some (small) ε corresponding to $\bar{\varepsilon}$.

2.5.2   Strategies and equilibrium

The game has essentially the same structure as the general stochastic signaling game studied in this chapter. The games differ in the flows and payoffs available for the two players. The description of the strategies and equilibrium concept carries over from Section 2.2.2. The entrant, given its belief π about the incumbent's type, chooses a strategy to enter at a sufficiently large market, that is whenever $x \geq U(\pi)$. The weak incumbent's strategy is to stop charging limit prices if the market variable x falls below $L(\pi)$. The second dimension of the incumbent's strategy is the choice of prices, in particular we need to specify the limit prices, i.e. the prices set in the pooling outcome. As in the simple one-shot games, in principle, a continuum of prices may be sustained in a pooling equilibrium. While we do not develop formal refinement criteria for the continuous-time game, we focus on the most plausible outcome, that is on efficient pooling at the prices of the strong incumbent. Let us denote the profit flow coefficient of the w-type charging the monopoly price of the s-type by M^p, where $M^p = M^w(C_1^s)$.

The payoffs of the firms are in flows of profit and a fixed cost of entry. In case of entry the weak incumbent faces a profit decline by factor $M^w - D_1^w$. The effective cost of signaling is now the difference between the profit flow with limit pricing and the unconstrained monopoly profit, that is $M^w - M^p$. If the entrant enters the market with belief π, it pays the entry cost K and its expected profit flow coefficient is $\hat{\pi}D_2^w$ where $\hat{\pi} = \pi + (1 - \pi)\varepsilon$. Putting these elements together we write the total expected payoff functions for both firms given a pair of strategies (L, U) and starting values (x, π) as follows:

$$
\begin{aligned}
R_1(x, \pi; L, U) &= \mathbb{E}_{x,\pi} \left[\int_0^{\tau_L \wedge \tau_U} e^{-rt} M^p X_t dt + \int_{\tau_U}^{\infty} e^{-rt} D_1^w X_t dt \mathbf{1}_{\tau_L \geq \tau_U} \right. \\
&\quad \left. + \left(\int_{\tau_L}^{\tau_w} e^{-rt} M^w X_t dt + \int_{\tau_w}^{\infty} e^{-rt} D_1^w X_t dt \right) \mathbf{1}_{\tau_L < \tau_U} \right], \\
R_2(x, \pi; L, U) &= \mathbb{E}_{x,\pi} \left[\left(\hat{\pi}_{\tau_U} \int_{\tau_U}^{\infty} e^{-rt} D_2^w X_t dt - K \right) \mathbf{1}_{\tau_L \geq \tau_U} \right. \\
&\quad \left. + \left(\int_{\tau_w}^{\infty} e^{-rt} D_2^w X_t dt - K \right) \mathbf{1}_{\tau_L < \tau_U} \right].
\end{aligned}
$$

The stopping times are defined analogously to Section 2.2.2. The discussion of Section 2.4.1 applies and in equilibrium the incumbent plays a mixed strategy, which is a continuous distribution function G over trigger strategies.

To develop the notation used in the equilibrium proposition below, let us now consider the terminal payoffs after the incumbent's type is revealed. The entrant's value, given that $\pi = 1$, is

$$W(x) = \left(\frac{x}{U(1)}\right)^{\beta_1} (d_2 U(1) - K),$$

where $d_2 = D_2^w/(r - \mu)$. The optimal entry trigger is

$$U(1) = \frac{\beta_1}{1 - \beta_1} \frac{K}{d_2}.$$

Given the entrant's strategy $U(1)$, the value of the incumbent of the w type is

$$\Omega(x) = \frac{M^w x}{r - \mu} - \left(\frac{x}{U(1)}\right)^{\beta_1} \frac{(M^w - D_1^w)U(1)}{r - \mu}.$$

Detailed derivations, that are similar to those in the baseline signaling game, and the proof of the equilibrium result in the following proposition are relegated to the appendix.

Proposition 2.10 *Let* $\hat{\Pi} = \Pi + (1 - \Pi)\varepsilon$ *and*

$$f_1(m, G, \tilde{U}) =$$

$$\frac{f_2(1 - \pi_0 G)^2[(1 - \beta_1)(1 - \beta_2)\hat{\Pi}d_2\tilde{U} - \beta_1\beta_2 K]\Delta_1}{(1 - \varepsilon)\pi_0(1 - \pi_0)d_2\tilde{U}^2(\Delta_1 - \Delta_2) - \pi_0(1 - \pi_0 G)\tilde{U}[(\beta_1 - \beta_2)W - \hat{\Pi}d_2\tilde{U}(\Delta_1 - \Delta_2) - K\Delta_2]},$$

$$f_2(m, G, \tilde{U}) = \frac{(\beta_2 - 1)(M^w - M^p)m\,(\delta_1 - \delta_2)}{(\beta_1 - \beta_2)\left[(M^w - M^p)\tilde{U}\delta_2 + (M^p - D_1^w)m\right]},$$

and denote by $(G^*, \tilde{U}^*)$ *the (unique) solution to the system of differential equations*

$$G'(m) = f_1(m, G, \tilde{U}), \quad G(L_1) = 1, \tag{2.32}$$

$$\tilde{U}'(m) = f_2(m, G, \tilde{U}), \quad \tilde{U}(L_1) = U(0) = \frac{\beta_1}{\beta_1 - 1}\frac{K}{\varepsilon d_2}. \tag{2.33}$$

Then a pair of strategies $(G, \tilde{U})$ *is the unique MPBE of the signaling game if*

$$G(m) = \begin{cases} 1 & \text{if } m < L_1, \\ G^*(m) & \text{if } L_1 \le m \le L_0, \\ 0 & \text{if } L_0 < m \le x_0, \ L_0 \ne x_0. \end{cases}$$

If $L_0 = x_0$, *then at* $t = 0$ *the incumbent of the w type reveals its type with probability* $G^*(L_0)$, *signals with probability* $1 - G^*(L_0)$, *while at* $t > 0$ *it plays according to G. The*

entrant contests at

$$\tilde{U}(m) = \begin{cases} \tilde{U}^*(L_1) & if \ m < L_1, \\ \tilde{U}^*(m) & if \ L_1 \leq m \leq L_0, \\ \tilde{U}^*(L_0) & if \ L_0 < m \leq x_0. \end{cases}$$

The lower bound on the support of G is given by the solution L_1 of

$$\{(\beta_1 - \beta_2)(r - \mu)\Omega(L_1) - [(\beta_1 - 1)M^w + (1 - \beta_2)M^p]m\}\left(\frac{U(0)}{L_1}\right)^{\beta_1}$$

$$+ (\beta_1 - 1)(M^w - M^p)m\left(\frac{U(0)}{L_1}\right)^{\beta_2} + (\beta_1 - \beta_2)(M^w - M^p)\tilde{U}(L_1) = 0 \quad (2.34)$$

and the upper bound is

$$L_0 = \min\{\bar{L}_0, x_0\}, \ \bar{L}_0 = \inf\{m \geq L_1 : G^*(m) = 0\}.$$

Finally, $U(1) = \frac{1}{d_2}K\beta_1/(\beta_1 - 1)$.

2.5.3   Implications

By introducing continuous time dynamics and uncertainty, we can derive some interesting implications that are unavailable in the existing game theoretic models of limit pricing. We formulate here several observations that are direct consequences of the equilibrium result and are of interest in the specific context of limit pricing as either empirical predictions or policy recommendations.

Observation 1 (Price dynamics) When the incumbent reveals its type at a random lower trigger, prices increase in a decreasing market.

Under our assumption of isoelastic demand and constant marginal cost, (unconstrained) monopoly prices are constant. Yet the price dynamics under limit pricing may take an unusual pattern with prices increasing in a decreasing market. (Under other demand-cost specifications, this would translate into an unexpected price increase in a decreasing market.) Firstly, this observation provides an empirical prediction that could be confronted with the data. Secondly, we provide a potential policy implication of this remarkable price dynamics. Based on one-shot models, limit (or predatory) pricing can be detected by comparing prices to marginal costs. This is the usual approach of antitrust authorities (see e.g. Sufrin and Jones (2004)). However, marginal costs are in general difficult to observe and thus in the regulatory practice undesirable limit pricing is difficult to discover and prove. The task to prove limit pricing practices is even more daunting when taking the asymmetric-information arguments into account,

with the assumption that costs are unobservable by competitors. The advantage of our dynamic model is that it implies that the easily observable price dynamics may reveal limit pricing practices of incumbents. In particular, increasing prices in a decreasing market indicate that the incumbent has used prices to deter entrants.

Observation 2 (Path dependence) The entrant's decision to enter depends on historical demand.

Our model shows that market dynamics (that is in our setup the transition from monopoly to duopoly) exhibits path dependence in that the entrant's decision to enter depends on historical demand. This is despite the fact that the demand shocks are Markovian and the current demand level is a sufficient statistic for the future distribution. Yet, because a market downturn in the past made it the more likely that the weaker type of incumbent would have stopped using limit pricing, the probability of facing the strong incumbent increases under the limit pricing regime. In other words, a demand slump polarizes entry timing, entry happens either early against the weak incumbent or late against an uncertain type.

Observation 3 (Learning and entry) The learning effect postpones entry.

Under complete information, the ratio of expected discounted profits at entry to the fixed cost is $\beta_1/(\beta_1 - 1)$ in both cases if $\pi = 0$ and $\pi = 1$ (recall that $\beta_1 > 1$, and the reason that the ratio exceeds 1 is that it incorporates the value of waiting, the standard result from the theory of investment under uncertainty, see, e.g., Dixit and Pindyck (1994)). Yet, when there is still incomplete information about the incumbent type, that is if $\pi \in (0, 1)$, the same ratio, that is $\hat{\pi} d_2 U(\pi)/K$, is larger than $\beta_1/(\beta_1 - 1)$. The difference stems from the learning effect. Whenever $\pi \in (0, 1)$, the entrant takes into account that over time it may learn more about the incumbent's cost type realization and make a more knowledgeable decision in the future. Consequently, the entrant postpones the entry decision and requires higher expected profits to enter.

2.6   Conclusions

We have presented a model of dynamic signaling in a stochastic environment and showed that such a setup brings novel strategic interactions between the informed and uninformed players. In our setting the payoffs (stake in the game) depend on the type of the informed player and follow a diffusion process. For a given belief about the type of the informed player, the uninformed player has incentives to stop the signaling game and contest the potential payoff at a sufficiently high stake. On the other hand,

the informed player has incentives to stop signaling at a sufficiently low stake. We characterize a Markov equilibrium in which the two players choose threshold strategies on the stake to stop the signaling game. Interestingly, the minimum process of the stake in the game captures the Bayesian learning of the uninformed player. Based on this observation, we could use the techniques of optimal stopping of extremum processes. The dynamic stochastic environment causes the gradual evolution from pooling via semi-separating to separating outcome.

The prospects of the model can be judged in the best way when our framework is applied to some concrete signaling situations. The dependence on the minimum process drives the path-dependence of the outcome of the game. In the limit pricing application, this feature brings a path-dependent market structure. Specifically, timing of entry into the market will depend on the past realizations of the demand. The model may be particularly valuable for applications in corporate finance. In corporate finance theory, both asymmetric information and continuous-time dynamics driven by diffusion processes play prominent roles. The stochastic signaling game merges these two, so far independent, modeling environments.

2.A   Appendix: Proofs

Proof of Proposition 2.1. For any $\mu \neq 0$, using that the solution to the differential equation for X_t if $\sigma = 0$ is $X_t = x_0 e^{\mu t}$, we derive the discount factor

$$e^{-rt} = \left(\frac{x_0}{X_t} \right)^{\frac{r}{\mu}}. \tag{2.35}$$

Suppose now that $\mu < 0$. Because x decreases deterministically, Player 2 enters whenever he breaks even in expectations. To see it, first note that any threat of Player 2 to enter earlier to induce type revelation of the w type is an empty threat. The w type would rather not reveal and make Player 2 believe that she is the s type. Second, for any given π there is no value in waiting as x decreases in time. So Player 2 enters whenever $X_t \geq U(\pi_t) = K/\pi_t$. If Player 1 reveals at some trigger above K, it spurs an immediate entry. Clearly, Player 1 obtains the highest payoff from signaling if she signals from $t = 0$ until the time when x reaches K and thus prevents any entry. The incentive compatibility constraint for such a signaling pattern is satisfied at $t = 0$ whenever

$$-\frac{c}{r} \left[1 - \left(\frac{x_0}{K} \right)^{\frac{r}{\mu}} \right] > -1. \tag{2.36}$$

The left-hand side represents the cost of signaling when x is between x_0 and K (using (2.35)), and the right-hand side represents the payoff if the type is revealed. It is easy to

verify that if the constraint (2.36) holds at x_0, it remains binding for any $x \in (K, x_0)$. So if $\Gamma_1(x_0) > 0$, Player 1 would only reveal whenever $X_t \leq L = K$. If $\Gamma_1(x_0) > 0$, the incentive compatibility constraint does not hold, and Player 1 does not signal at all, so $L = x_0$. Finally, when either $\Gamma_1(x_0) > 0$ or $\Gamma_1(x_0) < 0$, there does not exist an equilibrium where the w type randomizes between revealing and signaling as she strictly prefers one of the alternatives. ∎

Proof of Proposition 2.2. Player 2 cannot use a threat to enter early to induce type revelation of the w type. The w type would rather not reveal and make Player 2 believe that she is the s type. So for any π Player 2 chooses the level of the entry trigger in x to maximize its expected payoff $(x_0/x)^{r/\mu} (\pi x - K)$ (using (2.35)). This yields that Player 2 enters as soon as $X_t \geq U^d(\pi) = (K/\pi) r/(r - \mu)$.

Player 1 of the w type gains nothing from signaling if she reveals at $x \in (x_0, U(1)]$ and faces an immediate entry if she reveals at $x \in (U(1), U(\pi_0))$. It follows that, if Player 1 decides for signaling, the most profitable signaling strategy is to signal until Player 2 enters at $U(\pi_0)$. The incentive compatibility constraint for such a signaling pattern is satisfied at $t = 0$ whenever

$$-\frac{c}{r} - \left(\frac{x_0}{U(\pi_0)}\right)^{\frac{r}{\mu}} \left(1 - \frac{c}{r}\right) > -\left(\frac{x}{U(1)}\right)^{\frac{r}{\mu}}. \qquad (2.37)$$

The left-hand side represents the cost of signaling when x is between x_0 and $U(\pi_0)$ (using (2.35)), and the right-hand side represents the payoff if the type is revealed.

We now show that if the incentive compatibility constraint (2.37) holds at x_0 it will hold at any $x \in (x_0, U(\pi_0))$. At any $x \in [x_0, U(1)]$ the condition for signaling equivalent to (2.37) can be written as

$$-\frac{c}{r} + \left(\frac{x}{U(\pi_0)}\right)^{\frac{r}{\mu}} \left(-1 + \frac{c}{r}\right) + \left(\frac{x}{U(1)}\right)^{\frac{r}{\mu}}$$
$$\geq \left(\frac{x}{x_0}\right)^{\frac{r}{\mu}} \left[-\frac{c}{r} + \left(\frac{x_0}{U(\pi_0)}\right)^{\frac{r}{\mu}} \left(-1 + \frac{c}{r}\right) + \left(\frac{x_0}{U(1)}\right)^{\frac{r}{\mu}}\right] \geq 0,$$

where in the last inequality we use (2.37). Similar inequalities can be written for $x \in (U(1), U(\pi_0)]$. This proves that if (2.37) holds then Player 1 would never reveal until entry. So if $\Gamma_2(x_0) > 0$, Player 1 would never reveal, or, in terms of a trigger strategy, reveal whenever $X_t \leq L = 0$ (precisely, at any other inaccessible value in $(0, x_0)$). If $\Gamma_2(x_0) < 0$, pure strategy pooling is not an equilibrium. The (pure strategy) separating outcome is an equilibrium, if the weak type does not want to imitate the

strong type, that is if

$$-\frac{c}{r} \leq -\left(\frac{x}{U(1)}\right)^{\frac{r}{\mu}}. \tag{2.38}$$

The left-hand side represents the cost of signaling when Player 2 believes the type is s and never enters. It follows that if $\Gamma_3(x_0) \leq 0$, the unique equilibrium is a pure strategy of player to reveal at x_0. If neither (2.37) nor (2.38) holds, then the w type cannot put probability one on either pure strategy. The probabilities p and $(1-p)$ follow from Bayes rule and the indifference of Player 1 for signaling and revealing at $t = 0$. ∎

Proof of Proposition 2.10. The proof closely follows the logic of the arguments used in the general game. Here we concentrate on the points where some adjustments are needed. We begin with the derivation of the complete information payoffs, i.e. the terminal payoffs in the signaling game. If $\pi_t = 1$, then the entrant solves the optimal stopping problem

$$W(x) = \sup_{t \leq \tau \leq \infty} \mathbb{E}\left[\int_{\tau}^{\infty} e^{-r(u-t)} D_2^w X_u du - e^{-r(\tau-t)} K | X_t = x\right].$$

In the continuation region, i.e. for $x \in (0, U(1))$, $W(x)$ satisfies the following Bellman-type differential equation

$$rW(x) = \mu x W'(x) + \frac{1}{2}\sigma^2 x^2 W''(x) + D_2^w x$$

with the following three boundary conditions $W(U(1)) = d_2 U(1) - K$, $W'(U(1)) = d_2$ and $W(0) = 0$. Using these in the general solution to the differential equation yields $W(x)$ and $U(1)$ preceding the proposition. $U(0)$ can be found in a straightforward way. The value of the incumbent of w type $\Omega(x)$ satisfies a similar differential equation in $x \in (0, U(1))$, that is

$$r\Omega(x) = \mu x \Omega'(x) + \frac{1}{2}\sigma^2 x^2 \Omega''(x) + M^w x,$$

subject to $\Omega(U(1)) = D_1^w U(1)/(r - \mu)$ and $W(0) = 0$. The formula for $\Omega(x)$ given above the proposition follows.

Next, following Lemma 2.5, we characterize the condition on $U(m)$ such that $m \in \text{supp}(G)$. Let $F(x, m)$ be the value function of the incumbent of the w type satisfying the condition that the firm is indifferent between stopping signaling and continuing at $x = m$ for all $m \in \text{supp}(G)$. In the continuation region, for $x \in (m, \tilde{U}(m))$ with m

fixed, $F(x, m)$ must satisfy

$$rF(x, m) = \mu x F_x(x, m) + \frac{1}{2}\sigma^2 x^2 F_{xx}(x, m) - M^p x,$$

subject to the continuous and smooth fit conditions $F(\tilde{U}(m), m) = D_1^w \tilde{U}(m)/(r - \mu)$, $F(m, m) = \Omega(m)$, $F_x(m, m) = \Omega'(m)$, and the normal reflection condition $F_m(m, m) = 0$ for all $m \in \text{supp}(G)$. Solving the system of boundary conditions with the general solution we obtain

$$\left\{(\beta_1 - \beta_2)\,(r - \mu)\Omega(m) - [(\beta_1 - 1)\,M^w + (1 - \beta_2)M^p]\,m\right\}\left(\frac{\tilde{U}(m)}{m}\right)^{\beta_1}$$

$$+ (\beta_1 - 1)\,(M^w - M^p)\,m\left(\frac{\tilde{U}(m)}{m}\right)^{\beta_2} + (\beta_1 - \beta_2)\,(M^w - M^p)\,\tilde{U}(m) = 0.$$

After differentiating this implicit equation in $\tilde{U}(m)$ with respect to m we obtain (2.33). If in the same problem we use a boundary condition for $m = L_1$ at the upper trigger as $F(U(0), L_1) = D_1^w U(0)/(r - \mu)$, we obtain equation (2.34) defining L_1.

At the next step we derive the best response of the entrant to a continuous strategy G of the incumbent. Similar to Section 2.4.2 denote the value of the entrant in this best response problem by $V(x, m)$. $V(x, m)$ must satisfy the following differential equation

$$rV(x, m) = \mu x V_x(x, m) + \frac{1}{2}\sigma^2 x^2 V_{xx}(x, m).$$

The boundary conditions of continuous and smooth fit are $V(\tilde{U}(m), m) = \hat{\Pi}(m)d_2\tilde{U}(m) - K$ and $V_x(\tilde{U}(m), m) = \hat{\Pi}(m)d_2$. The boundary condition at the diagonal (m, m) is $V_m(m, m) = \Pi(m)g(m)\,(V(m, m) - W(m))$. After solving the set of boundary conditions with the solution to the differential equation and reorganizing we obtain (2.32) in the proposition, where, as before, $\Delta_1 = (m/\tilde{U}(m))^{\beta_1} - (m/\tilde{U}(m))^{\beta_2}$ and $\Delta_2 = \beta_2(m/\tilde{U}(m))^{\beta_1} - \beta_1(m/\tilde{U}(m))^{\beta_2}$.

The reminder of the proof is identical to the proof of Proposition 2.9. ∎

3

Corporate Liquidity and Solvency

3.1 Introduction

Financial distress is widely recognized as a key driving force behind corporate finance decisions. At the same time, however, the roles and interaction of liquidity and solvency distresses—two sources of financial distress—are not well understood. The main contribution of this chapter is the integration of liquidity and solvency concerns in a tractable intertemporal model of corporate finance with implications for valuation, capital structure, dividend policy, cash holdings, and credit spreads.

We build on the contingent claims models of risky asset valuation introduced by Black and Scholes (1973) and Merton (1974). Since Leland (1994), an important part of the literature has focused primarily on the corporate-finance implications of contingent claims modeling with the central role given to the optimal choice of capital structure. The standard structural trade-off model looks at the optimal choice of leverage that balances the tax benefits of debt and bankruptcy costs, and in which equity triggers default when the firm becomes insolvent. Subsequent extensions of this model has been successful in analyzing debt maturity, debt renegotiation, recapitalization, incomplete accounting information, macroeconomic regimes, debt structure and investment (Leland and Toft (1996), Fan and Sundaresan (2000), Goldstein, Ju and Leland (2001), Duffie and Lando (2001), Hackbarth, Miao and Morellec (2006), Broadie, Chernov and Sundaresan (2007), Hackbarth, Hennessy and Leland (2007), Sundaresan and Wang (2007)). This framework has been determinative for our understanding of credit risk and the role of debt in corporate finance.

Despite these developments, the structural trade-off framework has not been successful in treating at least three issues. First, the structural models typically predict

no role for corporate cash holdings, and thus they feature essentially zero cash. This omission is remarkable, especially in light of the recent surge of interest in corporate cash policies (Opler, Pinkowitz, Stulz and Williamson (1999), Almeida, Campello and Weisbach (2004), Faulkender and Wang (2006), Acharya, Davydenko and Strebulaev (2007)), which in turn has been driven by a significant increase in cash holdings among U.S. firms (Bates, Kahle and Stulz (2008)). Bates et al. (2008) report that an average U.S. firm holds 23.2% of its assets in cash—a significant share that remains unaccounted for by the trade-off theory. It is important to note that the empirical studies analyzing corporate cash and its increase frequently use the leverage ratio as one of the key regressors. However, it is intuitively clear that both variables are endogenous, with the debt level affecting the optimal cash level (via coupon payments, for example) and cash holdings influencing default risk and, thus, debt value. To gain a clear understanding of the relationship between cash and leverage, we need a structural model that endogenizes both capital structure and cash policy.

Second, the structural models treat dividends merely as balancing items. This leads to the unrealistic prediction that all residual cash flows are paid out to equity immediately. This trivial inactive dividend policy cannot provide any basis for empirical implications. It is certainly at odds with one of the most pervasive patterns of dividend payouts, namely, dividend smoothing (Lintner (1956), Brav et al. (2005)). Moreover, similarly to empirical studies on cash holdings, the empirical literature on dividends frequently uses leverage as a key independent variable, while we have no satisfactory theory that explains debt choice and endogenizes realistic (smooth) dividends.

Third, the standard trade-off models are also deficient in the way they treat default. Following Leland (1994), the standard approach is to study endogenous default that is triggered by equity holders when the firm becomes insolvent, i.e., when its equity value becomes negative. This approach excludes the other common reason for default, which is that, in the presence of financing constraints, the firm cannot cover its debt obligations due to liquidity distress. Davydenko (2007) reports that close to 10% of defaulting firms are economically solvent but face liquidity distress (with a caveat that the number might be underestimated due to default costs biasing asset values downwards) and concludes that "[neither solvency nor liquidity concerns] alone can fully explain observed default decisions."

In order to tackle these deficiencies we design a model in which both insolvency and illiquidity may cause default on debt payments. To construct a corporate environment suitable for our purpose, the model must have two characteristics. First, we must allow for financing constraints. Because a financially unconstrained firm will be able to raise external equity financing whenever it remains solvable (Leland (1994)), the assumption of financing constraints is a prerequisite to the relevance of liquidity

distress. Moreover, we suppose that the firm's cash flow process has *two* sources of uncertainty. First, instantaneous cash flows are subject to liquidity shocks and, second, expected instantaneous cash flow is uncertain. The idea is that the firm generates stochastic instantaneous cash flows that have a fixed mean depending on profitability of the firm, but the actual profitability is not directly observable. Investors form and update their expectations about the mean cash flow (or, in other words, about the firm's profitability) by using their prior belief about the distribution of cash flows and by learning from observing cash flow realizations. This learning process reflects the fact that persistent negative (positive) liquidity shocks translate into decreased (increased) expected profitability. The first source of uncertainty captures short-term liquidity distress. The second source of uncertainty is behind long-term financial distress, which eventually may lead to insolvency. We note that with cash flows characterized by uncertain profitability but without liquidity shocks, as modeled in the Leland's (1994) standard framework, instantaneous cash flows are predictable and liquidity management becomes trivial. In contrast, cash flows with liquidity shocks, but with a fixed expected flow, leave no room for solvency distress and solvency default.

As it turns out in our model, introducing financing constraints and liquidity concerns creates very plausible endogenous cash holding and dividend policies. Without cash reserves, the firm very soon becomes illiquid and is forced into default while still being solvent. This default is inefficient, as it would never have happened without financing constraints. We characterize a (variable) cash level, denoted by $\overline{M}$, that allows the firm to withstand liquidity shocks up to the point where the equity holders endogenously trigger solvency default. We show that $\overline{M}$ evolves over time and increases with expected profitability. Intuitively, a more profitable firm is more solvent and thus requires a larger cash reservoir to withstand more significant liquidity shocks before it is eventually declared insolvent. Consistent with empirical evidence, endogenous cash holdings serve as a buffer to absorb losses and as a means to avoid inefficient default (Opler et al. (1999), Lins, Servaes and Tufano (2007)).

We show that it is optimal for the firm that maximizes equity value to retain all earnings if cash is below $\overline{M}$ and, subsequently, to pay out dividends that allow the firm to maintain cash at (evolving) $\overline{M}$. The optimal dividend policy implied by our model is particularly notable. As in corporate practise, endogenous dividend flows are smooth in comparison with cash flows or earnings (preview Figure 3.3 for an illustration). The intuition behind the smoothing mechanism is the following. We note first that with a constant target level of cash reserves, dividend flows are tied to earning shocks. With cash at the target level, positive earnings are fully distributed and negative earnings lead to dividend omission. In our model, this is different. Suppose that the firm realizes surprising positive earnings. The firm that generates high cash flows is valued more

(the expectation of future instantaneous cash flows increases) and thus it requires more cash to cushion liquidity shocks before it becomes insolvent. Consequently, surprising positive earnings lead to an increase in optimal cash holdings, requiring more earnings to be retained instead of distributed to equity. In the case of surprising low earnings, the expectation of future cash flows decreases and so does the firm's valuation. Thus the firm becomes less solvent and the cash reserves needed to fend off liquidity distress before insolvency decrease. As a result, some cash is released and distributed to equity, complementing lower earnings. Both positive and negative earning shocks are smoothed out.

Another notable feature of our model is that the extension with liquidity concerns reduces the dispersion of the predicted credit spreads. This effect addresses the key problem with the predictive power of structural models as documented by Eom et al. (2004), whose empirical study indicates that in the case of relatively high spreads, the available structural models predict credit spreads that are too high, and credit spreads that are too low when the predicted spreads are relatively low. In other words, the predicted spreads are too dispersed. The reason that our model predicts less dispersed credit spreads is explained as follows. The financially constrained firm needs to raise the initial cash from external financing. We show that the exposure to liquidity distress and the firm's initial cash reserves are lowest for intermediate coupon levels. This is because there are two effects of coupon rates on optimal cash holdings. On the one hand, with increasing coupon rates, the firm's solvency risk is greater and the relative role of liquidity concerns decreases. This effect implies lower optimal cash reserves with higher coupon rates. But on the other hand, if coupon payments become relatively high, they impose a burden on cash flows and, thus, increase liquidity risk and optimal cash reserves. Consequently, the firm's cash needs are lowest for intermediate coupon payments. If external financing is subject to the proportional issuance cost, then the firm that minimizes this cost will tend to gravitate to the intermediate coupon levels that minimize the amount of needed cash. Therefore, with liquidity concerns and costly issuance, we observe less dispersion in the predicted optimal coupons, which translates into lower dispersion of credit spreads across firms.

Our analysis also indicates that short-term cash flow volatility and long-term uncertainty about a firm's economic prospects may have very different effects on financial variables. We show, for example, that cash holdings increase in short-term volatility and decrease in the magnitude of long-term uncertainty. Credit spreads decrease in short-term volatility, with the opposite effect is found with increasing long-term uncertainty. The two sources of uncertainty exhibit different effects because, essentially, short-term volatility is related to liquidity concerns, and long-term uncertainty to solvency concerns.

We find that the optimal leverage ratios in our model decrease relative to the benchmark model without liquidity concerns. This is a desirable feature because the standard model has been criticized for predicting excessive leverage. The main force behind this effect is the inclusion of endogenous cash holdings. Higher cash levels tend to decrease default risk and thus increase debt value. But it is the equity that is the direct claimant of cash, and thus marginal cash is directly accounted for in equity value. Consequently, cash holdings increase the denominator of the leverage ratio (firm value) more than its numerator (debt value).

In addition to the literature on contingent claims valuation of risky debt, this chapter relates to the literature on dynamic liquidity management and dividend payout optimization. Jeanblanc-Picqué and Shiryaev (1995) study a tractable model of a financially constrained firm threatened by costly liquidation, where the optimal payout policy is to retain all earnings if cash reserves are below a certain fixed threshold and to pay out everything otherwise. The model has been extended to incorporate, among others, investment and costly financing (Décamps and Villeneuve (2007), Løkka and Zervos (2008), Décamps, Mariotti, Rochet and Villeneuve (2007)). Remarkably, the literature on modeling dividends has so far not succeeded in producing a model of realistic dividends as observed in corporate practice, particularly, in demonstrating why firms smooth dividend payouts. Our model contributes to this literature by showing that adding uncertainty in the expected value of cash flow and concerns over solvency leads the optimizing firm to smooth dividends over cash flows.

Our work is also related to the recent work of DeMarzo and Sannikov (2007). In their model an agent controls the firm's expected cash flows of the firm through costly efforts, and the initially unknown expected profitability is learned over time. They show that the principle/investor can implement the optimal contract through a payout policy that is smoothed relative to cash flows. Both models, Demarzo and Sannikov's and ours, share the prediction of smooth dividends and the assumption of cash flows that are characterized by uncertain expectations and that are also subject to unpredictable shocks. The models differ, however, as Demarzo and Sannikov's results are built on the principle-agent conflict and focus mainly on payout policy, whereas our results follow from the trade-off arguments. Moreover, our model covers a broader area of corporate finance beyond payout policy (debt coupon, taxes, bankruptcy cost, flotation cost) and, while building on the standard contingent claims analysis, may be more suitable for the valuation of corporate securities and the analysis of credit risk. When taken together, the two models imply that the two sources of uncertainty in cash flows may produce smooth dividends in different modeling setups.

Several recent papers also feature both cash holdings and debt financing. Hennessy and Whited (2005) presents a trade-off model in which firms use a mix of equity,

one-period debt and cash balance to cover their financing needs. Their model includes refinancing and endogenous investment, the features absent in our model. On the other hand, in Hennessy and Whited (2005) firms never hold both debt or positive cash balance at the same time, so in effect cash is simply treated as negative debt. In their model default is precluded which results in riskless debt and zero credit spreads. Our model produces formulas for valuation of risky debt and implications for credit spreads. Moreover, our analysis is focused on the roles of short-term liquidity and long-term solvency distresses, while the framework of Hennessy and Whited (2005) does not model and distinguish these forces. A newer paper by Gamba and Triantis (2008) extends Hennessy and Whited (2005) and allows firms to hold both debt and cash holdings at the same time, but the other differences remain. Acharya, Almeida and Campello (2007) recognize that, as in our paper, the presence of financing frictions is a precondition for a meaningful role of cash holdings in corporate policy. In comparison to our paper, their motivation for cash is different and is based on the distinct roles of cash and negative debt in hedging future investment opportunities against future cash flows. The scope of the analysis is more limited than in our paper, where we also study implications for payout policy and credit risk. Acharya, Huang, Subrahmanyam and Sundaram (2006) introduce cash holdings into a discrete-time model of risky debt and solve it numerically. In contrast to our model, their focus is on the role of strategic debt renegotiation. In general, our analysis with closed-form results is more tractable than previous models that relied on numerical solutions (except for the simple model of Acharya, Almeida and Campello (2007)).

In the following section, we set up the model, then in Section 3.3, we analyze a benchmark case of a firm without financing constraints, concerned only about solvency. Section 3.4 presents the main model with both liquidity and solvency concerns. In Section 3.5 we discuss the impact of liquidity concerns on corporate finance and derive a set of empirical predictions. Section 3.6 presents our conclusions, and the Appendix provides the proofs omitted in the main text.

3.2   Setup

3.2.1   Outline and timing

We assume that management behaves in the interest of equity holders, all investors are risk neutral, and discount cash flows at a constant risk-free rate r. The model is set in continuous time with an infinite horizon; time is indexed as $t \in [0, \infty)$.

The original equity holders are financially constrained and seek external financing to cover investment cost I and initial cash reserves M_0. Investment cannot be delayed.

Once successfully financed, the firm generates a continuous flow of earnings, with cumulative earnings at time t denoted as X_t. The earnings process is the main state variable and is described in detail in the next subsection. Earnings are subject to corporate taxes at rate τ with full loss offset provision. The debt coupon payments are deducted from earnings for tax purposes, creating the tax benefit of debt. Corporate cash reserves earn interests at the risk-free rate r.

The financing may come from a combination of equity and perpetual debt that promises flow coupon c. The value function of equity is denoted E and that of debt is D. We allow for both fixed and proportional flotation costs of new issuance, denoted $L \geq 0$ and $\lambda \in [0, 1)$, respectively. For the sake of simplicity, the costs are the same for both debt and equity.

The sequence of events and decisions is as follows. At time $t = 0$ the firm issues a combination of equity and debt to maximize the value of the original equity holders. After that the firm starts receiving the flow of earnings and pays out the promised coupon and corporate taxes. Net profits (or losses) are left at the disposal of the firm and are either retained to increase (decrease) cash reserves or are paid out to equity holders as dividends (in the case of instantaneous losses dividends may be paid out from positive cash reserves). Cumulative dividends up to time t are denoted by Div_t.[1] To deal effectively with indeterminate situations, we assume that equity holders pay out marginal cash holdings whenever they *weakly* prefer to do so.

When the firm has no means to cover the current coupon payments it defaults for the reasons of illiquidity. We call such an event a *liquidity default*. The financial distress is driven here by short-term factors. The firm may also, acting in the interest of equity holders, voluntarily default if the value of equity falls below zero. In this case, the firm is not profitable enough for the equity holders to run it and pay the debt coupons. Then, the firm faces long-term distress; we refer to this type of default as a *solvency default*.

In the event of either type of default, the firm is liquidated, which is costly. The debt claims have the absolute priority in the case of default and the liquidation value is αA, $\alpha \in (0, 1)$. Here $1 - \alpha$ is the proportional liquidation cost and A is the value of the all-equity firm at the moment of default.[2]

---

[1] We choose to refer to the payouts to equity holders as dividends but the whole model applies to stock repurchases as well.

[2] Following the standard in the literature, we simplify the analysis by assuming that the firm is not refinanced with an optimal capital structure after default.

3.2.2   Earnings and uncertainty

The firm generates a stochastic flow of earnings before interest and taxes (EBIT):

$$dX_t = \overline{\mu}dt + \sigma d\overline{Z}_t, \tag{3.1}$$

where $\overline{\mu}$ is the mean of EBIT, σ is its volatility and $\overline{Z}$ is a standard Brownian motion. All parties (insiders and outsiders) have the same information at each time t. They observe the cumulative EBIT process $\{X_s, s \leq t\}$ that generates a filtration $\{\mathcal{F}_t\}$. There are two sources of uncertainty. First, instantaneous flows are subject to Brownian shocks $d\overline{Z}_t$, which represent short-term liquidity shocks. Second, the profitability of the firm is uncertain, which is represented by the fact that the true mean $\overline{\mu}$ is ex ante unknown to all parties. We assume that $\overline{\mu}$ is fixed and can take either of the two values μ_L or μ_H, with $\mu_L < \mu_H$. All parties share a common prior expectation μ_0 about $\overline{\mu}$, with $\mu_0 \in (\mu_L, \mu_H)$.

The two sources of uncertainty serve to capture the two main sides of corporate financial distress. The unpredictable immediate earnings (due to Brownian shocks) bring in the short-term liquidity risk. The uncertain drift $\overline{\mu}$ puts the firm in a position to undergo solvency distress and, ultimately, solvency default.

As time evolves, more information becomes available and the parties update their expectation of mean earnings. The current set of information generated by X_t is described by $\mathcal{F}_t$ and is used in a Bayesian fashion to update the conditional expectation to

$$\mu_t = \mathbb{E}\left[\overline{\mu}|\mathcal{F}_t\right].$$

We can use the optimal filtering theory to find the law of motion of the posterior expectation variable. Let us introduce an innovation process Z as the difference between the realized and expected earnings, defined by the differential equation

$$dX_t = \mu_t dt + \sigma dZ_t. \tag{3.2}$$

The process Z is a Brownian motion adapted to filtration $\mathcal{F}_t$. Note that Z differs from $\overline{Z}$ (which is not observable by the parties and not adapted to $\mathcal{F}_t$). Equation (3.2) describes the dynamics of X in terms of observables.

A version of Theorem 9.1 in Liptser and Shiryaev (2001) then yields that the posterior expectation of the mean earnings level evolves as

$$d\mu_t = \frac{1}{\sigma}\left(\mu_t - \mu_L\right)\left(\mu_H - \mu_t\right)dZ_t. \tag{3.3}$$

Note first that the posterior expectation process is a martingale as it incorporates all predictable information. Second, the volatility of μ is inversely related to σ, reflecting the fact that expectations adjust more rapidly if the noise term in the earnings process is small (the earnings signals are informative). Finally, learning slows down as evidence accumulates in favor of one state and μ is close to either μ_L or μ_H.

3.2.3 Relation to existing literature

The main framework of our model closely follows the standard in the literature on contingent claims modeling of capital structure based on the trade-off theory. The distinguishing feature of our model is the specification of the cash flow process in (3.1), which, with the use of filtering theory, can be rewritten as (3.2) and (3.3).

The motivation for our modeling choice is three-fold with the first two reasons stemming from the need to expand and connect the two areas of literature related to our analysis. First, cash flows in our specification are subject to unpredictable liquidity shocks to introduce non-trivial cash and dividend policy. This is similar to in liquidity management models that analyze optimal dividend policy and predict precautionary cash reserves that cushion liquidity shocks (Jeanblanc-Picqué and Shiryaev (1995)). Technically, cumulative cash flows are modeled here as a stochastic process following an arithmetic Brownian motion. As a result, instantaneous cash flows are *increments* of the process and are subject to Brownian shocks.[3] In contrast, the structural default literature typically models instantaneous cash flows as the *level* of a geometric Brownian motion, in which case, cash flows are predictable and liquidity management becomes trivial.

Second, we also allow for the drift of the arithmetic Brownian motion to be uncertain to enable endogenous solvency default. In the models based on a simple arithmetic Brownian motion with constant drift, the expected profitability is constant, and, given fixed debt obligations, the firm is always either solvent or insolvent, erasing the endogenous default from the model. With our assumption of uncertain drift, the firm may become insolvent, in the sense that it is not profitable enough for equity holders to cover its debt obligations (as in Leland (1994), Leland and Toft (1996) and others).

Third, it is analytically convenient to assume cash flows following the stochastic differential equation (3.1). Specifically, we obtain closed-form solutions for corporate securities values, optimal cash reserves, dividends and a default threshold. The same stochastic environment has been successfully adapted in different contexts by Moscarini

---

[3]Instantaneous cash flows have also been modeled as increments of an arithmetic Brownian motion in the continuous-time agency-based models of corporate finance (DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin and Rochet (2007)).

(2005) to study job matching in labor markets and Keppo, Moscarini, and Smith (2008) to analyze the value of and demand for information.

3.3   Solvency default without liquidity concerns

For the sake of comparison we start with a benchmark. Following the framework introduced by Leland (1994) we assume in this section that the firm is not subject to liquidity default. The endogenous solvency default is triggered by equity holders when equity value becomes negative. The equity holders are willing and able to inject any funds necessary to keep operations running whenever the equity value is positive. Following Leland (1994), secondary equity financing proceeds are not subject to flotation costs. As in numerous contingent claims models of capital structure, a closed-form solution is available under the simplifying assumption that debt is issued only once at the initial date (Leland (1994), Leland and Toft (1996), Fan and Sundaresan (2000), Duffie and Lando (2001), Miao (2005), Hackbarth et al. (2007), Sundaresan and Wang (2007)).[4] Accordingly, we assume the following.

Assumption 3.1 *New debt financing is constrained to time $t = 0$.*

Assumption 3.2 *Equity financing is costless beyond $t = 0$.*

Under these assumptions the firm is without liquidity concerns and there is no room for cash holdings because any liquidity needs can be covered by an injection of equity financing. We use subscript u with the value functions in this section to denote the financially unconstrained case. For brevity, we suppress the dependence of the value functions on other parameters except for μ, but most notably they also depend on coupon c.

We first consider the value of the firm if it were financed fully by equity. If we assume that $\mu_L \geq 0$,[5] then the firm is always profitable and its value is simply equal to the expected discounted future after-tax cash flows:

$$A_u(\mu) = \mathbb{E}\left[\int_0^\infty e^{-rt}(1-\tau)\,dX_t|\mu_0 = \mu\right] = (1-\tau)\frac{\mu}{r}.$$

The liquidation value that debt holders receive in the event of default is $\alpha A_u(\mu)$, with $1 - \alpha$ representing the proportional liquidation cost.

---

[4]In an alternative and more complex setup, Goldstein et al. (2001) allow for upward leverage adjustments.

[5]The alternative assumption that $\mu_L < 0$ would introduce an optimal liquidation of the firm even in the absence of debt financing. In this case, $A_u(\mu)$ equals the expected discounted future after-tax cash flows up to the time of liquidation, which is optimally chosen by the equity holders. We omit this minor extension, which adds little to our model, while slightly raising the complexity of expressions.

At the next step we find the values of the claims held by the debt and equity holders. These values depend on the flows to the claimants and the default time. The optimal default time, chosen by the equity holders, is the first time expected profitability μ falls to some threshold μ_u^*.

The firm issues perpetual debt that pays a constant continuous coupon at rate c per unit of time. It follows from the standard arguments and Itô's lemma that, before default, debt value D_u satisfies the following ordinary Bellman-type differential equation:

$$rD_u(\mu) = \frac{1}{2\sigma^2} (\mu - \mu_L)^2 (\mu_H - \mu)^2 D_u''(\mu) + c, \tag{3.4}$$

subject to

$$D_u(\mu_u^*) = \alpha A_u(\mu_u^*), \ D_u(\mu_H) = \frac{c}{r}.$$

This system states that if the firm is not in default, the required rate of return on the debt equals the sum of the coupon flow and the expected increase in the value of debt. At μ_u^* the firm defaults and the debt is valued at $\alpha A_u(\mu_u^*)$. The boundary condition at μ_H, which is an absorbing state for μ, asserts that D_u is bounded and equal to the risk-free value.

At each period t before default, the equity receives the expected flow of $(1 - \tau)(\mu_t - c)$, which is the expected free cash flow after taxes and coupon payments. As in general $\mu_u^* < c$ (confirmed below in (3.8)), this means that non-negative dividends are expected as long as $\mu_t \geq c$ and that in periods with $\mu_t < c$, equity receives "negative dividends" in expectation. The negative distributions are typically interpreted in this type of models as equity issuances. This implies that, unrealistically and inconsistently with evidence on costly equity issuance, the firm resorts to frequent external, especially when close to default. We address this issue in our main model in Section 3.4, below.

Within this setting, the equity value E_u must satisfy the following differential equation:

$$rE_u(\mu) = \frac{1}{2\sigma^2} (\mu - \mu_L)^2 (\mu_H - \mu)^2 E_u''(\mu) + (1 - \tau)(\mu - c), \tag{3.5}$$

subject to

$$E_u(\mu_u^*) = 0, \ E_u(\mu_H) = (1 - \tau)\frac{\mu - c}{r}.$$

This equation and the boundary conditions can be interpreted similarly to the ones for debt valuation.

Having defined equity and debt values, we can calculate total levered firm value F_u, which, by definition, equals the sum of equity and debt:

$$F_u(\mu) = E_u(\mu) + D_u(\mu). \tag{3.6}$$

The equity holders choose the default trigger ex post—after the initial financing. This means that they maximize equity value E_u over μ_u^*, which is equivalent to setting the smooth pasting condition on $E_u(\mu)$ at μ_u^*:

$$E_u'(\mu_u^*) = 0. \tag{3.7}$$

The condition requires the optimal value function to be smooth at the default trigger, and indeed, it can be shown that it corresponds to the first order condition from maximization of $E_u(\mu)$ with respect to μ_u^*.

The optimal capital structure is determined at the issuance point with the choice of coupon c, which maximizes the value of the initial equity holders (to indicate the dependence on c directly, we add it as a parameter to the value functions in the remainder of this section). The firm seeks to finance the investment cost I with debt and new equity. If the new equity holders obtain a fraction ϕ of the equity and if the proportional and fixed issuance costs are λ and L then the following financing identity holds

$$I = (1 - \lambda)\left(D_u(\mu_0, c) + \phi E_u(\mu_0, c)\right) - L,$$

which can be rewritten as

$$(1 - \phi)\, E_u(\mu_0, c) = D_u(\mu_0, c) + E_u(\mu_0, c) - \frac{I + L}{1 - \lambda}.$$

The left-hand side represents the value of the initial equity holders. Hence, maximization of the left-hand side is equivalent to maximization of $E_u(\mu, c) + D_u(\mu, c)$. It then follows, using (3.6), that the optimal choice of coupon c (and thus of the initial leverage) by the initial equity holders is equivalent to maximizing of $F_u(\mu_0, c)$.

We summarize the findings of this section in the following proposition.

Proposition 3.1 *Suppose Assumptions 3.1 and 3.2 hold and $\mu_L \geq 0$. The optimal solvency default is characterized by the first time μ is at or below μ_u^* given by*

$$\mu_u^* = \frac{\mu_L \mu_H + \left[(\beta - 1)\,\mu_H - \beta \mu_L\right] c}{(1 - \beta)\,\mu_L + \beta \mu_H - c}. \tag{3.8}$$

If $\mu \geq \mu_u^*$, the values of equity $E_u(\mu)$, debt $D_u(\mu)$ and total firm $F_u(\mu)$ are given by

$$E_u(\mu) = (1 - \tau)\frac{\mu - c}{r} - \left(\frac{\mu - \mu_L}{\mu_u^* - \mu_L}\right)^{1-\beta}\left(\frac{\mu_H - \mu}{\mu_H - \mu_u^*}\right)^{\beta}(1 - \tau)\frac{\mu^* - c}{r}, \tag{3.9}$$

$$D_u(\mu) = \frac{c}{r} + \left(\frac{\mu - \mu_L}{\mu_u^* - \mu_L}\right)^{1-\beta}\left(\frac{\mu_H - \mu}{\mu_H - \mu_u^*}\right)^{\beta}\left(\alpha A_u(\mu_u^*) - \frac{c}{r}\right), \tag{3.10}$$

and

$$F_u(\mu) = (1-\tau)\frac{\mu}{r} + \tau\frac{c}{r} - \left(\frac{\mu-\mu_L}{\mu_u^*-\mu_L}\right)^{1-\beta}\left(\frac{\mu_H-\mu}{\mu_H-\mu_u^*}\right)^{\beta}\left[(1-\alpha)\,A_u(\mu_u^*) + \tau\frac{c}{r}\right], \quad (3.11)$$

where

$$\beta = \frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{8r\sigma^2}{(\mu_H-\mu_L)^2}} > 1. \qquad (3.12)$$

The optimal coupon rate c^* maximizes $F_u(\mu_0)$ over c.

The closed-form expressions for the value functions are interpreted as follows. The value of equity (3.9) is the sum of the present value of perpetual distributions to equity and the present value of cash flows lost at default. The value of risky debt in (3.10) consists of two terms. The first term, c/r, is the value of risk-free perpetual debt. The second term reflects the impact of default risk and equals the present value of cash flows lost by debt in case of default. Total firm value (3.11) consists of three elements: the first one is the present value of the perpetual flow of net earnings, the second is the present value of the tax benefits of debt, and finally, the negative term corrects for the present value of the cash flows lost at default.

Equation (3.8) implies that, in general, $\mu_u^* < c$ (see also the discussion below Proposition 3.5 and Figure 3.2). This means that, as in other structural default models following Leland (1994), the equity holders expect negative cash flows when close to default, yet they prefer to keep the firm running. Moreover, it is worth noting that neither the proportional flotation cost λ nor the fixed one L influences the optimal choice of c.

3.4   Model with liquidity concerns

Following the standard in the related literature, in the previous section we assumed that, after the initial issuance, equity could be issued frequently and without cost and that the debt flotation costs (or other implicit concerns) would prohibit debt re-issuance. Empirical evidence clearly indicates the opposite: the issuance costs of debt, both fix and variable costs, are significantly lower than those of equity (Altinkiliç and Hansen (2000), Leary and Roberts (2005)). Leary and Roberts (2005) further document that new equity is issued less frequently than debt. Clearly, the assumption of the benchmark model, that new equity serves to cover current coupon payments in the case of insufficient earnings, is difficult to reconcile with this evidence.

To address this issue in a tractable way, we restrict the firm's access to external financing. After the initial issuance, which is subject to fixed and proportional costs,

the firm cannot raise additional capital. This simplifying assumption, which facilitates the analysis with closed-form solutions, can be justified by the fixed issuance cost and also by the same convention that excludes secondary debt issuance in numerous contingent claims models of capital structure (see the references above Assumption 3.1). For further reference we introduce the following assumption.

Assumption 3.3 *New external financing is constrained to time $t = 0$.*

As in the benchmark case, debt holders' claims have absolute priority over the productive assets in the case of default. However, the firm now also holds liquid non-productive assets, namely cash reserves, and we assume that these are distributed to equity just before default. We abstract from any possible contracts that might limit such distributions as they are not central to our model.[6] This assumption simplifies the analysis and, moreover, as we show below, in most cases the optimizing firm reaches the endogenous trigger with zero cash holdings.

3.4.1   Cash and dividend policy

At each time before default the firm generates stochastic EBIT dX_t and pays out tax-deductible debt coupon cdt. The dynamics of earnings net of taxes and debt obligations, denoted by Y_t, is thus

$$dY_t = (1 - \tau)(dX_t - cdt) = (1 - \tau)(\mu_t - c)\,dt + (1 - \tau)\,\sigma dZ_t. \qquad (3.13)$$

Without cash reserves and with financing constraints, the firm becomes illiquid and is forced into default as soon as $dX_t < cdt$. In our model, positive cash reserves serve as a means to decrease liquidity risk. Let us denote cash reserves at time t by M_t. Cash reserves change at each time by the instantaneous interest earned on current cash holdings and the difference between net earnings and dividend payout:

$$dM_t = rM_t dt + dY_t - dDiv_t. \qquad (3.14)$$

In general, the higher M_t, the lower is the risk of liquidity distress. Of special interest is the level of cash holdings that allows the firm to avoid liquidity default altogether. The next proposition characterizes this level of cash reserves.

---

[6] In any case, such covenants may be difficult to enforce as the equity holders would try to preempt with the cash distributions.

Proposition 3.2 *Let $\overline{M}$ be the lowest level of cash reserves that allows the firm to avoid liquidity default under Assumption 3.3. $\overline{M}(\mu)$ is given by*

$$\overline{M}(\mu) = (1-\tau)\left[\frac{\sigma^2}{\mu_H - \mu_L}\ln\left(\frac{\mu - \mu_L}{\mu_H - \mu}\frac{\mu_H - \mu^*}{\mu^* - \mu_L}\right) + \max\left\{0, \frac{1}{r}\left(c - \frac{\mu_L + \mu_H}{2}\right)\right\}\right].$$
(3.15)

The proof, given in the Appendix, relies on the requirement that the dividend process Div_t is non-decreasing. This requirement implies a set of ordinary differential equations, with (3.15) being the minimal solution satisfying these equations.[7]

Before interpreting the expression for $\overline{M}$ in (3.15), we first determine the dividend stream that is implied by the cash policy $M_t = \overline{M}(\mu_t)$. First, by Itô's lemma the dynamics of $\overline{M}$ is

$$d\overline{M}_t = (1-\tau)\left[\mu_t - \frac{1}{2}(\mu_H + \mu_L)\right]dt + (1-\tau)\sigma dZ_t.$$
(3.17)

Then using (3.14) and (3.17), if $M_t = \overline{M}(\mu_t)$ the dividend stream is given by

$$dDiv_t = r\overline{M}_t dt + dY_t - d\overline{M}_t = \left[r\overline{M}_t + (1-\tau)\left(\frac{\mu_H + \mu_L}{2} - c\right)\right]dt.$$
(3.18)

Note that with (3.15), we can write $dDiv_t$ as a function of μ_t only and not directly of $\overline{M}_t$.

Equation (3.15) implies that $\overline{M}$ increases with μ and decreases with μ^*. This is because, if the current μ is closer to default at μ^*, the liquidity shocks to be absorbed by cash before endogenous solvency default become smaller. The effect of the coupon rate on cash holdings is twofold. The main effect works for all levels of c indirectly via μ^*. A higher c means earlier default or, equivalently, higher μ^* (see (3.29) below) and thus lower $\overline{M}$. The direct effect results from the last term of (3.15). Thus, it

---

[7]An alternative and instructive way to see the result is to think of M_t as the level of cash that is sufficient to withstand a shock in Z_t that brings μ_t to μ^* (irrespective of how quickly the shock is realized). For brevity, we focus here on the case of $c \leq \frac{1}{2}(\mu_H + \mu_L)$. Equation (3.14) then implies that $\overline{M}(\mu_t) = (1-\tau)\sigma(Z_t - Z^*)$, where $Z_t - Z^*$ is the shock that brings μ_t to default trigger μ^*. To characterize $Z_t - Z^*$, let us define $\theta_t = f(\mu_t) = \frac{\sigma^2}{\mu_H - \mu_L}\ln\frac{\mu_t - \mu_L}{\mu_H - \mu_t}$ and $\theta^* = f(\mu^*)$ (note that $\theta_t = \theta^*$ if and only if $\mu_t = \mu^*$). Applying Ito's lemma to θ_t, we have

$$\theta_{t'} = \theta_t + \int_t^{t'}\frac{1}{2\sigma^2}(2\mu_s - \mu_H - \mu_L)ds + \frac{1}{\sigma}(Z_{t'} - Z_t).$$

This equation also holds for $\theta_{t'} = \theta^*$ in particular. So the shock that brings θ_t to θ^* (and also μ_t to μ^*) is $Z_t - Z^* = \sigma(\theta_t - \theta^*)$. It follows that $\overline{M}(\mu_t)$ must satisfy

$$\overline{M}(\mu_t) = (1-\tau)\sigma(f(\mu_t) - f(\mu^*)) = (1-\tau)\frac{\sigma^2}{\mu_H - \mu_L}\ln\left(\frac{\mu - \mu_L}{\mu_H - \mu}\frac{\mu_H - \mu^*}{\mu^* - \mu_L}\right),$$
(3.16)

which confirms (3.15) in the proposition for the case $c \leq \frac{1}{2}(\mu_H + \mu_L)$. To obtain the additional term in (3.15), one must impose the condition that the implied dividend payout is not negative for all $\mu_t > \mu^*$ (which is not the case under (3.16) if $c > \frac{1}{2}(\mu_H + \mu_L)$).
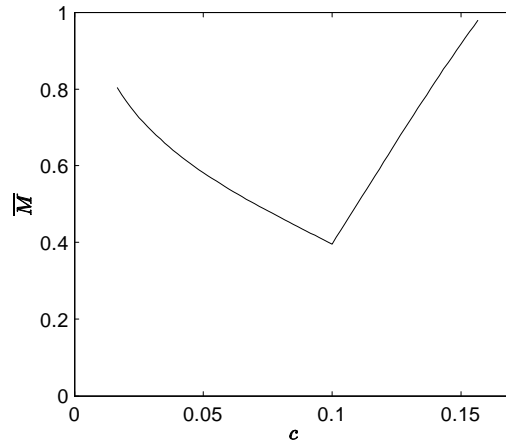
FIGURE 3.1. Target cash $\overline{M}$ as a function of coupon c. The parameter values are: $\mu_L = 0$, $\mu_H = 0.2$, $\sigma = 0.2$, $r = 0.06$, $\tau = 0.15$, $\alpha = 0.6$, and $\mu_0 = 0.1$.

works only if $c > \frac{1}{2}(\mu_H + \mu_L)$ and results in $\overline{M}$ increasing in c. If c is high relative to the expected profitability, then higher cash holdings are required to complement the operational cash flows in meeting high debt obligations. Figure 3.1 illustrates the effects of the coupon on $\overline{M}$ (the parameter values are as calibrated in Section 3.5). The total effect of changes in the coupon is such that cash holdings decrease in c for small c and increase if c exceeds $\frac{1}{2}(\mu_H + \mu_L)$.

As $\overline{M}$ is the lowest level of cash reserves that allows the firm to avoid liquidity default, it is not surprising that $\overline{M} = 0$ as μ reaches μ^* in case c is not too large $(c \leq \frac{1}{2}(\mu_H + \mu_L))$. If c is larger than $\frac{1}{2}(\mu_H + \mu_L)$, then high coupon payments require positive cash holdings at all times before default. Note that the additional term in (3.15) when $c > \frac{1}{2}(\mu_H + \mu_L)$, that is $\frac{1}{r}\left[c - \frac{1}{2}(\mu_H + \mu_L)\right]$, makes the dividend rate in (3.18) equal to zero at default.

Suppose that the dividend-cash policy aims at decreasing the risk of liquidity default. We later verify that this is indeed optimal if the firm's objective is to maximize equity value. Intuitively, this suggests that all cash flows are retained if the firm is at risk of liquidity default and that dividends are paid out as long as such distributions do not bring in liquidity risk. To characterize this proposed dividend policy more formally, let us denote it by Div_t^* at each time t. If, for a given μ_t, the cash reserves are below the target level $\overline{M}$, the firm retains all the earnings:

$$dDiv_t^* = 0 \text{ if } M_t < \overline{M}(\mu_t). \tag{3.19}$$

If the cash level is at $\overline{M}(\mu_t)$, the payout policy is such that this level is maintained as μ_t fluctuates. This is, according to (3.18):

$$dDiv_t^* = \left[rM_t + (1 - \tau)\left(\frac{\mu_H + \mu_L}{2} - c\right)\right]dt \text{ if } M_t = \overline{M}(\mu_t). \quad (3.20)$$

If the cash level exceeds $\overline{M}(\mu_t)$, the residual is paid out:

$$dDiv_t^* = M_t - \overline{M}(\mu_t) \text{ if } M_t > \overline{M}(\mu_t). \quad (3.21)$$

Before proving that this cash-dividend policy is optimal for equity holders, we demonstrate an intuitive property of optimal equity value that states that the partial derivative of the optimal equity value $E(\mu, M)$ with respect to M is larger than or equal to one. This is intuitive because any extra cash holdings can be paid out immediately as dividends, and the optimal dividend policy followed again. To see it, note that for any cash level M, equity value $E(\mu, M)$ of the firm following the optimal dividend policy must be at least equal to the sum of optimal equity value with $M - \Delta M$ cash, $E(\mu, M - \Delta M)$, and ΔM in a dividend payout: $E(\mu, M) \geq E(\mu, M - \Delta M) + \Delta M$. After rearranging the inequality and letting ΔM go to zero, we obtain

$$E_M(\mu, M) \geq 1. \quad (3.22)$$

We can state the following about the dividend policy.

Proposition 3.3 *The payout policy (3.19)-(3.21) maximizes equity value.*

Intuitively, the proposed payout policy is optimal because it directs the retention of all cash flows whenever marginal cash holdings decrease the probability of illiquidity (so that the cash withheld in the firm is worth more than its face value, $E_M(\mu, M) > 1$) and the payout of excess cash flows otherwise (when marginal cash holdings in the firm are equal to their face value, $E_M(\mu, M) = 1$).

3.4.2   Valuation of corporate securities

The values of corporate securities depend on a large number of factors, among them the initial cash level financed by external investors. To obtain closed-form solutions, we assume that the firm issues securities sufficient to cover cash holdings $\overline{M}(\mu_0)$, which allow the firm to avoid liquidity risk.

Assumption 3.4 $M_0 = \overline{M}(\mu_0)$.

We note that this assumption is partially validated by Assumption 3.3, which constrains the availability of external financing to the initial date. Without additional external financing, all the required cash is raised with the initial issuance.[8] If $M_0 = \overline{M}(\mu_0)$, then by Proposition 3.3 the optimal dividend policy is given in (3.20) for all $\mu_t > \mu^*$. This payout policy implies that $M_t = \overline{M}(\mu_t)$ for all $\mu_t > \mu^*$. In other words, under Assumption 3.4, the firm holds cash reserves at the level $\overline{M}(\mu_t)$ until the endogenous solvency default and is hedged against liquidity risk.

Under our assumptions, debt value D equals the present value of continuous coupon payments up to the time of default as soon as μ_t reaches μ^*. $D(\mu)$ must satisfy the following differential equation:

$$rD(\mu) = \frac{1}{2\sigma^2} (\mu - \mu_L)^2 (\mu_H - \mu)^2 D''(\mu) + c.$$

At default debt holders receive a fraction α of the EBIT-generating technology. That is, following the earlier literature, we simplify the financing issues after default. This implies that the debt holders recover $\alpha A(\mu^*)$ at default, where $A(\mu) = \alpha (1 - \tau) \mu / r$ if $\mu_L \geq 0$. Thus, the differential equation for D is coupled with the following boundary conditions:

$$D(\mu^*) = \alpha A(\mu^*), \ \ D(\mu_H) = \frac{c}{r}.$$

With the assumptions of the present model, up to default at the first time μ_t falls to μ^*, the equity receives a flow of dividends equal to

$$dDiv_t = a_1 \ln \left(\frac{\mu - \mu_L}{\mu_H - \mu} \frac{\mu_H - \mu^*}{\mu^* - \mu_L} \right) dt + a_2 dt,$$

where

$$a_1 = \frac{(1 - \tau) r \sigma^2}{\mu_H - \mu_L}$$

and

$$a_2 = (1 - \tau) \max \left\{ 0, \left(\frac{\mu_L + \mu_H}{2} - c \right) \right\}.$$

Then it follows from the standard arguments that equity value E must satisfy the ordinary differential equation:

$$rE(\mu) = \frac{1}{2\sigma^2} (\mu - \mu_L)^2 (\mu_H - \mu)^2 E''(\mu) + r\overline{M}(\mu) + a_1 \ln \left(\frac{\mu - \mu_L}{\mu_H - \mu} \frac{\mu_H - \mu^*}{\mu^* - \mu_L} \right) + a_2,$$
$$(3.23)$$

---

[8] Note that with a sufficiently high variable issuance cost λ, the firm might prefer issuing securities for less than $I + \overline{M}(\mu_0)$ (but more than I) and collecting the remaining cash up to $\overline{M}$ from the retained earnings. The firm would balance the cost of exposure to liquidity risk and the benefit of cheaper source of capital. We assume this possibility away to obtain closed-form solutions.

subject to the following boundary conditions:

$$E(\mu^*) = \overline{M}(\mu^*), \ \ E(\mu_H) = (1 - \tau)\frac{\mu_H - c}{r} + \overline{M}(\mu_H). \tag{3.24}$$

As usual, the left-hand side of (3.23) reflects the required rate of return per unit of time for holding equity. The right-hand side represents the expected change in equity value plus the dividend flow per unit of time. The boundary condition at μ^* is in line with the assumption that the equity holders withdraw non-productive liquid assets prior to default. The boundary condition at μ_H ensures that $E(\mu_H) - \overline{M}(\mu_H)$ is bounded and equal to the risk-free value of free cash flows.

Solving the respective differential equations with the boundary conditions, we obtain closed-form solutions for both equity and debt values. The following proposition shows these results.

Proposition 3.4 *Suppose Assumptions 3.3 and 3.4 hold. Then for a given μ^* and $\mu \geq \mu^*$ debt and equity value satisfy*

$$D(\mu) = \frac{c}{r} + \left(\frac{\mu - \mu_L}{\mu^* - \mu_L}\right)^{1-\beta} \left(\frac{\mu_H - \mu}{\mu_H - \mu^*}\right)^{\beta} \left(\alpha A(\mu^*) - \frac{c}{r}\right), \tag{3.25}$$

and

$$E(\mu) = \overline{M}(\mu) + (1 - \tau)\frac{\mu - c}{r} - \left(\frac{\mu - \mu_L}{\mu^* - \mu_L}\right)^{1-\beta} \left(\frac{\mu_H - \mu}{\mu_H - \mu^*}\right)^{\beta} (1 - \tau)\frac{\mu^* - c}{r}, \tag{3.26}$$

with β given in (3.12).

Equation (3.25) implies that, for a given coupon c and default trigger μ^*, the debt value is the same as in the benchmark case reported in equation (3.10). This is not surprising as the liquidity risk in the present model is effectively hedged by appropriate cash holdings. Nevertheless, liquidity concerns may affect the optimal coupon and default and thus indirectly alter debt value.

Combining (3.9) and (3.26) reveals that equity value equals $E_u(\mu) + \overline{M}(\mu)$, which is the equity value of the firm without the liquidity constraints given in (3.9) plus the cash stock. By holding cash reserves $\overline{M}(\mu)$, the firm is hedged against liquidity distress and thus the value of its productive assets is equal to those of the financially unconstrained firm. In addition, equity holders hold the full rights to the cash holdings (which they nevertheless prefer to retain in the firm) and thus $\overline{M}(\mu)$ augments their value. The cash in the firm $\overline{M}(\mu)$ is worth exactly $\overline{M}(\mu)$ to the equity holders because the interest gained on cash equals the investors' discount rate.

By definition, the levered firm value equals the sum of the equity value and the debt value. From Proposition 3.4, we obtain that if $\mu_L \geq 0$, it holds that

$$F(\mu) = E(\mu) + D(\mu) = \overline{M}(\mu) + (1 - \tau)\frac{\mu}{r}$$

$$+ \tau\frac{c}{r} - \left(\frac{\mu - \mu_L}{\mu^* - \mu_L}\right)^{1-\beta} \left(\frac{\mu_H - \mu}{\mu_H - \mu^*}\right)^{\beta} \left[(1 - \alpha)(1 - \tau)\frac{\mu^*}{r} + \tau\frac{c}{r}\right]. \qquad (3.27)$$

Equation (3.27) demonstrates that the firm value is a sum of four components. It consists of the face value of cash holdings plus the present value of earnings net of taxes plus the present value of tax shield of debt minus the probability-adjusted present value of cash flows lost at default. Using (3.6), $F(\mu)$ can be written as $F_u(\mu) + \overline{M}(\mu)$. That is, the levered firm with liquidity concerns and with cash holdings that hedge liquidity risk equals the value of the firm without liquidity concerns plus the face value of the cash.

3.4.3   Default and optimal capital structure

Under Assumptions 3.3 and 3.4, the firm uses cash reserves to cushion liquidity shocks. Then the timing of default is endogenously selected by the equity holders. Default takes place at the moment that the firm is not solvent enough. The default policy takes the form of a lower threshold on μ, which maximizes equity value. This is achieved at μ^*, which satisfies the smooth pasting condition:

$$E'(\mu^*) = \overline{M}'(\mu^*). \qquad (3.28)$$

(Compare it with the smooth pasting condition (3.7) and the boundary condition for E at $\mu = \mu^*$ in (3.24) in the present model.)

The initial equity holders using new equity and debt seek to finance the investment cost I and the initial level of cash reserves $\overline{M}(\mu_0, c)$ (to stress the dependence on c, we add parameter c to cash and value functions in the rest of this section). If the new equity holders obtain a fraction ϕ of equity and if the proportional cost of issuance of both debt and equity is λ and the fixed cost of issuance is L, then the following financing identity holds:

$$I + \overline{M}(\mu_0, c) = (1 - \lambda)(D(\mu_0, c) + \phi E(\mu_0, c)) - L.$$

This can be rewritten as

$$(1 - \phi)E(\mu_0, c) = D(\mu_0, c) + E(\mu_0, c) - \frac{\overline{M}(\mu_0, c)}{1 - \lambda} - \frac{I + L}{1 - \lambda}.$$
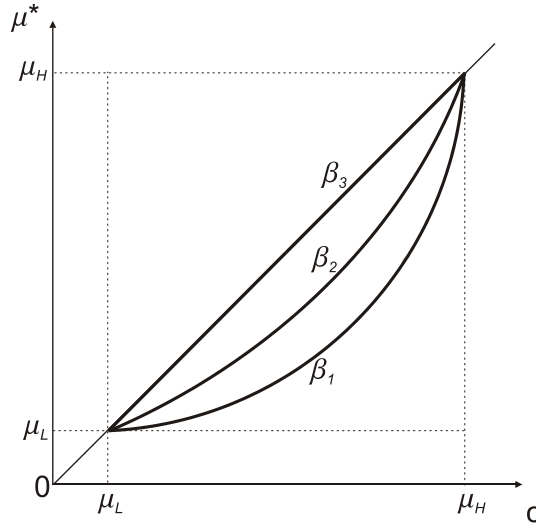
FIGURE 3.2. Optimal default trigger μ^* as a function of debt coupon c for various values of β with $\beta_1 < \beta_2 < \beta_3$ and $\beta_3 \to \infty$. β increases in σ and r and decreases in $(\mu_H - \mu_L)$.

The left-hand side represents the value to the initial equity holders. It follows that the optimal c that maximizes $(1 - \phi) E(\mu_0, c)$, also maximizes the right-hand side, and the objective function can be expressed as (3.30) in the next proposition. In the same proposition we also present the solution to the smooth pasting condition (3.28) for the optimal default trigger.

Proposition 3.5 *Under Assumptions 3.3 and 3.4 the optimal solvency default is characterized by the first time that μ is at or below μ^*, given by*

$$\mu^* = \frac{\mu_L \mu_H + [(\beta - 1) \mu_H - \beta \mu_L] c}{(1 - \beta) \mu_L + \beta \mu_H - c}. \tag{3.29}$$

The optimal coupon rate c^* maximizes

$$F(\mu_0, c) - \frac{\overline{M}(\mu_0, c)}{1 - \lambda} \tag{3.30}$$

over c.

Figure 3.2 presents the main properties of the optimal default trigger function (3.29). μ^* is a convex increasing function of c. It is intuitive that μ^* is equal to μ_H (μ_L) with coupon equal to μ_H (μ_L). This is because, with $c = \mu_H$, the equity holders expect losses for all μ and thus default immediately with $\mu^* = \mu_H$. When $c = \mu_L$, the firm generates positive expected profit net of coupon for all μ except at the absorbing state at μ_L, and thus the equity value is maximized with a default at $\mu^* = \mu_L$. For the intermediate values of c in (μ_L, μ_H), the default threshold falls below the coupon rate; in the figure, μ^* lies below the diagonal $\mu^* = c$. This difference between the expected

earnings at default and coupon represents the value of waiting to default. Because of this value, the equity holders prefer to keep the firm running despite that the coupon obligations exceed the expected earnings.

As illustrated in Figure 3.2, default triggers μ^* increase in β. By Equation (3.12), β depends on the earnings signal quality (that is on σ and on $\mu_H - \mu_L$) and the discount rate. It follows that the default trigger increases with the noisiness of the earnings signals (higher σ or smaller $\mu_H - \mu_L$) and with the level of discount rate r. Intuitively, with noisy signals and high r, the value of postponing default in order to wait for new information decreases.

μ^* in equation (3.29) is the same as μ_u^* in the benchmark case reported in (3.8). Since the firm is effectively hedged against liquidity distress, it makes sense that the solvency default trigger that maximizes equity value is the same as for the financially unconstrained firm. Interestingly, this is despite the precautionary cash reserves that need to be held in the firm. However, the isomorphism of μ^* and μ_u^* means only that the default policy in both cases is the same if coupon obligations are the same. The second part of Proposition 3.5 implies that in general the optimal coupons differ in the two cases with and without liquidity concerns.

Using (3.27), the objective function (3.30) can be rewritten as

$$F_u(\mu_0, c) - \frac{\lambda}{1 - \lambda} \overline{M}(\mu_0, c). \tag{3.31}$$

Comparing this objective function with the one of the financially unconstrained firm (which was $F_u(\mu_0, c)$), we note the major difference between the cases. Whereas the coupon choice in the benchmark analysis was independent of any issuance cost, the optimal coupon of the constrained firm is dependent on the proportional issuance cost λ. This is because now the capital structure choice interferes with the firm's financing needs: the firm needs to raise capital to cover the initial cash holdings, and the required initial level of cash depends on the coupon rate itself. As raising additional units of cash is costly due to the variable issuance cost, the firm's optimal choice of c also takes into account its impact on the initial amount of cash to be raised. Recall from Figure 3.1 that $\overline{M}$ is decreasing in c for low levels of c and increasing for high c. It follows that to minimize the flotation cost of raising the initial cash reserves, the constrained firm issues more debt than the unconstrained firm if the unconstrained firm's optimal coupon is relatively low (below $(\mu_L + \mu_H)/2$). The opposite happens if the unconstrained firm's optimal coupon is high (above $(\mu_L + \mu_H)/2$).

We note that, in the absence of financing frictions in the sense of zero variable cost of issuance ($\lambda = 0$), the objective function simplifies to $F_u(\mu_0, c)$ and is exactly equivalent

to the problem in the case without liquidity constraints. Moreover, the fixed cost of issuance does not matter for the choice of the optimal c.

3.5   Impact of liquidity concerns on corporate finance

In this section we analyze the implications of our model with liquidity concerns along two dimensions. First, we discuss the main differences between the standard trade-off model and our model with both solvency and liquidity concerns. Second, we examine a set of empirical implications of the extended model.

Changes in exogenous parameters typically affect a number or all endogenous variables simultaneously. We analyze the comparative statics implied by our model using the base case as a reference level. The base case parameter values are the following: $\mu_L = 0$, $\mu_H = 0.2$, $\sigma = 0.2$, $r = 0.06$, $\tau = 0.15$, $\alpha = 0.6$, $\lambda = 0.1$, and $\mu_0 = \frac{1}{2}(\mu_H + \mu_L) = 0.1$. The initial value of the expected cash flows is the mean of the binomial distribution. The volatility of cash flows is chosen such that the (initial) coefficient of variation (that is, σ/μ_0) is equal to 2. This corresponds to the annualized coefficients of variations reported in Irvine and Pontiff (2008)—they are equal to 1.59 for cash flows and 2.42 for earnings. Our choice of the proportional flotation cost of $\lambda = 0.1$ is above the parameter values estimated in some other studies (Gomes (2001), Hennessy and Whited (2005)), and is justified by our focus on firms that are fully financially constrained beyond the initial issuance. The values of the risk free rate r, the tax advantage of debt τ, and the recovery rate α closely correspond to the recent calibration exercises; see, for example, Hackbarth et al. (2006).

3.5.1   Cash holdings

The structural models of capital structure and credit risk following Leland (1994) have typically assumed away a meaningful cash policy. As in our benchmark analysis in Section 3.3, the equity holders are assumed to have no financial constraints and equity issuance is costless. Consequently, any necessary funds are provided by new equity issuance as long as the equity holders are willing to continue operating the firm. This leaves the cash policy irrelevant.

In contrast, our model predicts a non-trivial role for cash holdings. The firm holds a positive amount of cash to meet debt coupon payments in case these obligations exceed current earnings. In other words, with costly external financing, cash reserves serve as a cushion to prevent short-term liquidity distress. Our model further specifies that cash reserves are not meant to cover any losses. If the firm persistently generates losses for a longer time period, the (expected) profitability decreases and, ultimately,

the firm becomes insolvent. As a result, the optimal policy prescribes cash holdings that are a function of the expected earnings and are sufficient to cover liquidity shocks up to the point of endogenous default (the target level of cash is given by equation (3.15)).[9]

It is worth noting that the cash ratio (defined as cash holdings divided by total firm value) that is implied by our model is in line with cash holdings observed among U.S. firms. With our base case parameters, the cash ratio equals 20.6%. This value is similar to the average cash ratio of 23.2% documented for a sample of U.S. firms in 2006 by Bates et al. (2008).

The model predicts that cash holdings of financially constrained firms are strongly correlated with cash flows (compare (3.17) and (3.2)), while cash holdings of unconstrained firms are not systematically related to cash flows. This implication provides an alternative interpretation of the evidence of Almeida et al. (2004) that shows the same pattern of cash flow sensitivity of cash holdings. Almeida et al. (2004) explain their findings and precautionary cash holdings by the firms' need to fund future investments while facing financing constraints. In contrast, in our fully dynamic model, a constrained firm uses positive cash flows to build up cash (and uses cash to cover negative cash flows) in order to avoid inefficient default in the future. Our interpretation seems more in line with the empirical evidence in the study of Opler et al. (1999) that shows cash holdings as serving mainly to cover losses (and not capital expenditures or payouts to equity holders).

The empirical literature has been interested in the impact of debt on corporate cash holdings, treating the former variable as exogenous. Figure 3.1 presents the cash level $\overline{M}$ as a function of coupon c and shows that cash decreases in debt for low and moderate levels of debt and increases with high levels of debt. The empirical evidence of Opler et al. (1999) documents a negative relationship between cash and leverage. A more refined study by Guney, Ozkan and Ozkan (2007) provides evidence for a similar non-monotonic relationship between cash holdings and debt. Our model implies that such correlations may be expected from the data but it also suggests caution when interpreting the evidence and inferring any causal relationship since both cash stock

---

[9] The characteristics of the optimal cash policy predicted by our model seem to closely reflect corporate practice. Based on an extensive survey among international CFOs, Lins et al. (2007) conclude:

> [S]trategic cash serves a basic function—to provide a general purpose buffer against future cash shortfalls. CFOs state that this is the primary driver of strategic cash holdings—with its importance ranking far exceeding that of other response choices. Thus, it appears that firms use strategic cash to insure against all types of negative shocks to cash flows, rather than to just fund growth when external capital may not be available. This finding positions strategic cash holdings as a form of financial distress (or bankruptcy) insurance.

and debt level are endogenous and variations are driven by changes in exogenous variables.

Another relationship of interest is the impact of cash flow volatility on cash holdings. Equation (3.15) for the target cash stock of the constrained firm reveals that, ceteris paribus, higher EBIT volatility σ induces higher optimal cash holdings. In our model, this is because larger cash reserves are required to hedge against more volatile cash flows. This prediction is consistent with the empirical findings of Opler et al. (1999) and Han and Qiu (2007).

Our model predicts that the marginal value of cash holdings to equity holders differs across firms. In particular, the model is able to encompass all the main hypotheses of the recent empirical study of Faulkender and Wang (2006). Recall that the marginal value of cash is equal to 1 for both unconstrained firms and constrained firms at or above the target cash level $\overline{M}$. The marginal value of cash in constrained firms with cash below $\overline{M}$ exceeds one, reflecting the decreasing probability of liquidity default with an additional unit of cash. It is clear that the marginal value of cash is larger for constrained firms and that it decreases with the level of cash holdings. Most interestingly, we can derive a clear interpretation of the negative cross-sectional relationship between the marginal value of cash and debt level documented by Faulkender and Wang (2006) (Faulkender and Wang (2006) seem to build their hypothesis and interpretation on the contingent claims models, which do not have a meaningful cash policy). For small and moderate levels of debt, the comparative static exercise presented in Figure 3.1 reveals that the target level of cash decreases in debt. Then, for a fixed cash level below $\overline{M}$, an increase in debt implies that the current cash holdings are closer to $\overline{M}$ so the firm is closer to being fully hedged against liquidity shocks. Consequently, the marginal value of cash decreases in debt. Our model also predicts an untested possibility that the relationship is reversed for high levels of debt.

3.5.2 Smooth dividends

The standard structural trade-off models of capital structure treat dividends simply as a balancing item: any residual cash flows are paid out to equity holders. This leads to a dividend pattern that bears little resemblance to actual corporate payout decisions. As in our benchmark case in Section 3.3, in these models the implied payouts in each period constitute 100% of positive free cash flows, and dividends are omitted in periods of negative free cash flows.

Our model, being extended by liquidity concerns, predicts a very different optimal payout policy. When cash reserves are at the target level $\overline{M}(\mu)$, the level that prevents liquidity default, the optimal dividend payout is given by (3.20). This dividend payout
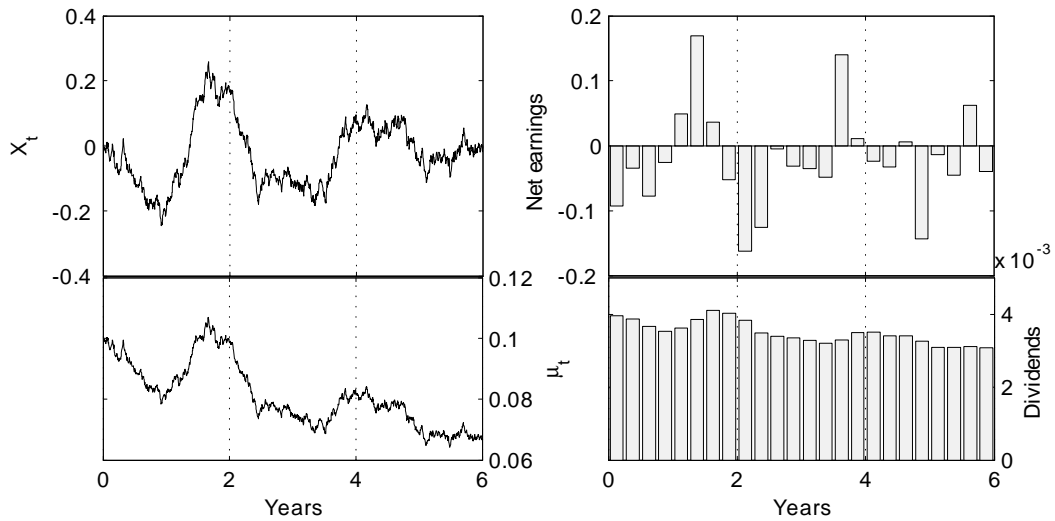
FIGURE 3.3. Simulated quarterly earnings and dividends. The parameter values are $\mu_L = 0$, $\mu_H = 0.2$, $\sigma = 0.2$, $r = 0.06$, $\tau = 0.15$, $\alpha = 0.6$, $\lambda = 0.1$ and $\mu_0 = 0.1$.

allows the firm to maintain cash reserves at $\overline{M}(\mu)$ with changing μ. From (3.20) we observe that in each period, dividend payouts are not only non-negative for $\mu \geq \mu^*$, which is by nature the case for corporate dividends, but also that the instantaneous payout is predictable. The latter fact is due to the absence of the Brownian shock in (3.20). This is in contrast to the dynamics of net earnings in (3.13), which apart from the time drift component, also include a Brownian motion term. This implies that net earnings are more volatile than dividends. In other words, the model predicts that dividends are smoothed relative to earnings. This prediction is in line with persistent evidence on the corporate practice of dividend payouts (Lintner (1956), Brav et al. (2005)).

Figure 3.3 illustrates the dividend smoothing produced by the model. The left-hand panel presents a simulation of EBIT process X_t and posterior expectations μ_t. We then use the model with liquidity concerns of Section 3.4 to calculate optimal coupon, default and cash holdings. The right-hand panel shows quarterly net earnings and dividends from this simulation. Clearly, the net earnings are positive and negative in different quarters, but these changes are only partly reflected in dividend changes. The dividends remain relatively stable. Even in the case of losses, the firm continues to pay out dividends.

This smoothing feature is driven by the interplay of liquidity and solvency default and the role of cash holdings as a cushion against liquidity shocks. The mechanism can be described as follows. Positive earnings shocks that bring in disposable cash flows also increase expected profitability. A more profitable firm is more valuable and thus it requires more cash reserves to fend off liquidity distress before declaring solvency
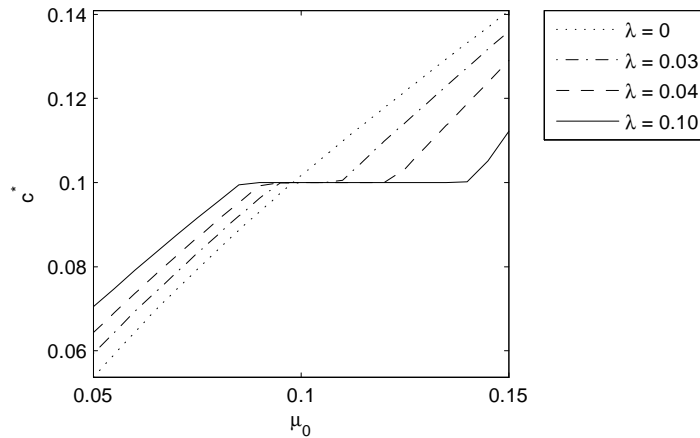
FIGURE 3.4. The effect of variable issuance cost λ on the choice of optimal coupon c^* for different initial profitability μ_0. The other parameter values are $\mu_L = 0$, $\mu_H = 0.2$, $\sigma = 0.2$ $r = 0.06$, $\tau = 0.15$, $\alpha = 0.6$.

default. In other words, dividends are flattened in the case of high earnings because an increase in cash flows is offset by increasing optimal cash reserves. In the case of surprisingly low earnings, expected profitability decreases, the firm gets closer to endogenous solvency default, and the cash level that allows it to avoid liquidity distress decreases. Consequently, low earnings lead to a release of some of the cash holdings that are paid out to equity. Both positive and negative earnings surprises are smoothed out, and as Figure 3.3 demonstrates, our model predicts positive and stable dividends even if earnings are very volatile.

Another feature of the endogenous dividend policy that is supported by empirical evidence is the prediction that firms in distress will rather reduce dividends but not omit them. This is documented by DeAngelo and DeAngelo (1990).

3.5.3   Issuance costs

Propositions 3.1 and 3.5 imply that issuance costs have no role in the choice of the optimal capital structure of the unconstrained firm (at least if the costs are uniform for all types of financing), and that the proportional issuance cost λ affects the optimal capital structure in the case of financing constraints and liquidity concerns. As explained before, λ matters because the funds to be raised from external investors— that is the investment cost and the initial cash $I + \overline{M}$—depend on the structure of the financing via $\overline{M}$.

The role of λ in determining the optimal coupon and thus the optimal capital structure is illustrated in Figure 3.4. We plot four curves of optimal coupon c^* for varying initial level of expected EBIT μ_0; each curve represents a different level of λ. The dotted line depicts the case of $\lambda = 0$. As usual, higher cash flows allow the firm to take more

debt and c^* strictly increases in μ_0. With a positive λ, the firm takes into account how much cash it requires to hedge liquidity risk. As $\overline{M}$ is the lowest at $c = \frac{1}{2}(\mu_H + \mu_L)$ (see Figure 3.1), minimizing the total issuance cost causes the optimal choice of c to gravitate towards $\frac{1}{2}(\mu_H + \mu_L)$. Consequently, we observe that for the unconstrained firm, for relatively low (or high) coupon payments, financing constraints increase (decrease) the optimal coupon. The effect is stronger with higher λ. This example reveals that a very small issuance cost may have a notable effect on the optimal coupon and capital structure. A realistic cost of $\lambda = 0.1$ leads to a pronounced distortion in the optimal coupon.

The proportional issuance cost has important implications for corporate credit spreads. Credit spreads are defined by the difference between the debt yield and the risk-free rate, $c/D - r$. Because, in the case of financing constraints and issuance cost, the optimal coupons are flattened, we may expect these factors to contribute to a decreased dispersion of credit spreads when compared to the financially unconstrained case. This effect allows us to address the key problem with the predictive power of structural models as reported by Eom et al. (2004). Eom et al. (2004) test the yield spread predictions of several structural models and conclude that the available models tend to produce too high a dispersion of predicted credit spreads. Where the structural models predict high credit spreads, these predictions notably exceed the actual spreads, and where the models predict low credit spreads these predictions fall significantly below the observed ones. Our model, with liquidity concerns, moves the predicted credit spreads in the desired direction. We illustrate this decreased dispersion of credit spreads in the following two subsections when we calculate the spreads for various parameter values.

3.5.4   EBIT volatility

Increasing EBIT volatility σ has two main direct effects on the endogenous variables. First, it increases the magnitude of liquidity shocks and, thus, liquidity risk. Second, it makes the instantaneous cash flows less informative about the true profitability $\overline{\mu}$. Less informative signals lead to an increase in μ^* due to a lower value of waiting with the decision to default (see Figure 3.2).

Figure 3.5 presents the effects of changes in σ. The displayed values are calculated with default triggers and coupons at the optimal level for each σ. The solid line plots the respective values for the financially constrained firm with liquidity concerns, and the dashed line plots the values for the unconstrained firm. Figure 3.5.A reveals that the issuance cost λ has a significant effect on the optimal coupon when liquidity concerns matter. The unconstrained firm reduces the optimal coupon c_u^* in increasing σ, because of higher default risk. The situation of the constrained firm is different, as it also has

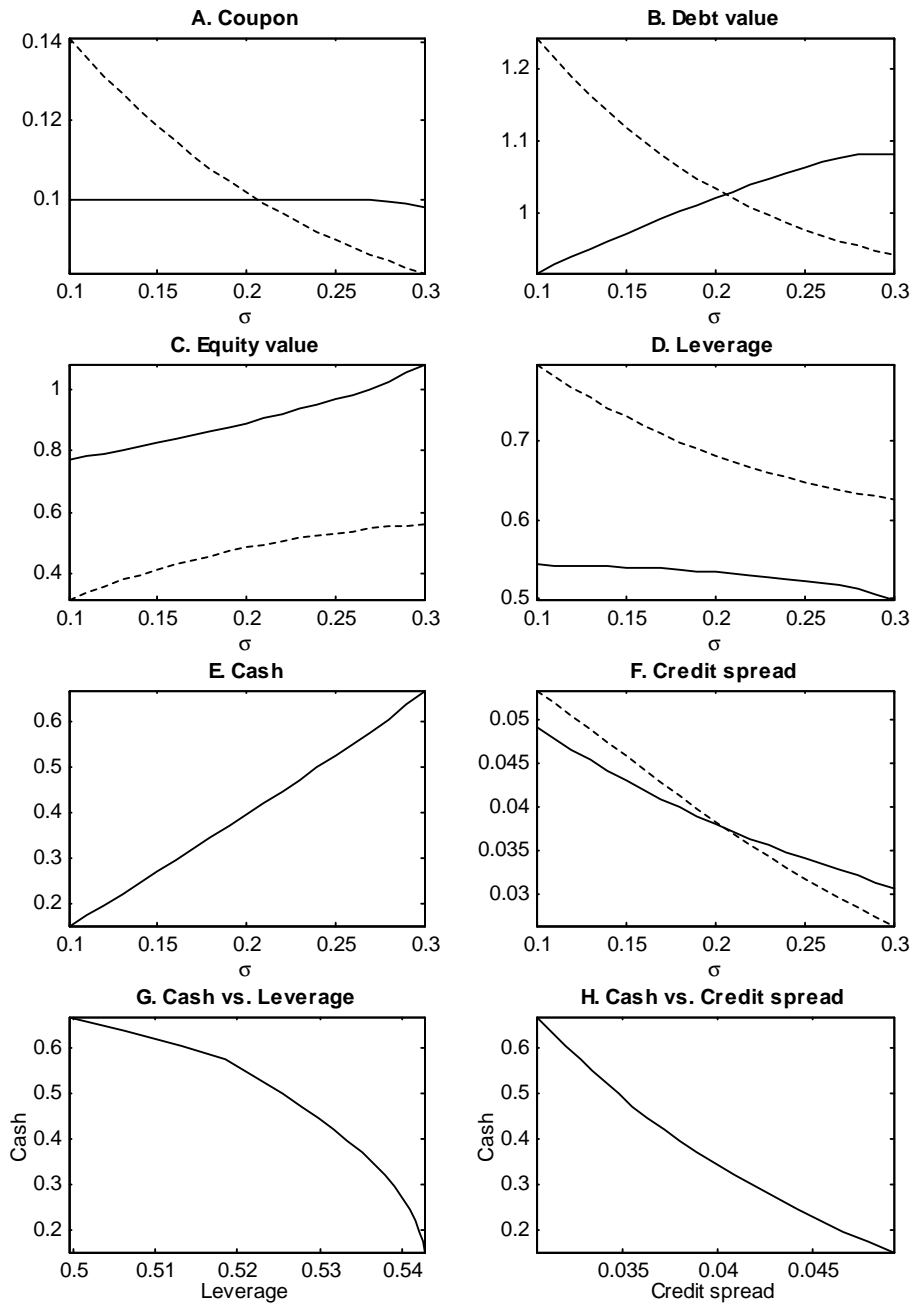FIGURE 3.5. Effects of EBIT volatility σ. The solid lines plot the respective values of the financially constrained firm with liquidity concerns, and the dashed lines plot the values of the unconstrained firm. Default and leverage are determined endogenously. The other parameter values are $\mu_L = 0$, $\mu_H = 0.2$, $r = 0.06$, $\tau = 0.15$, $\alpha = 0.6$, $\lambda = 0.1$ and $\mu_0 = 0.1$.

to raise funds for the initial cash reserves, and the target cash level $\overline{M}$ is the lowest for intermediate coupon rates (see Figure 3.1). Hence, while minimizing the issuance cost, the firm's optimal coupon c^* is driven towards the mean (see also Figure 3.4 and Section 3.5.3).

The behavior of the optimal coupon explains also the difference in debt values for the constrained and unconstrained firms (Figure 3.5.B). Whereas the unconstrained firm's debt value decreases in σ, the opposite occurs for the firm without liquidity concerns. It shows that in the case with liquidity concerns the positive effect of maintaining a relatively high and stable coupon dominates other effects that increase the default trigger and decrease debt value.

The plot in Figure 3.5.D, showing the leverage ratio, reveals that, despite the differences in debt values, the leverage ratio decreases in both cases. This is in accordance with the empirical evidence on leverage (Titman and Wessels (1988)). In the case of the constrained firm, the decrease in leverage is the result of the increase in equity value exceeding the decrease in debt value (see Figure 3.5.C). The forces that push the equity value up in σ are, first, the well-known call-option characteristics of equity (equity benefits from positive shocks and has an option to default in case of negative shocks) and, second, the increase in cash holdings. Due to a larger liquidity risk in increasing cash flow volatility, optimal cash holdings increase in σ, as shown in Figure 3.5.E. This prediction is consistent with the empirical evidence documented by Bates et al. (2008). Our analysis confirms that the explanation in Bates et al. (2008), that the recent spectacular expansion in cash holdings among U.S. firms is to a large degree due to the increasing riskiness of cash flows, has a theoretical grounding in a model with endogenous cash and financing.

Figure 3.5.F shows that in the case of financing constraints and issuance costs, the predicted credit spreads are less dispersed than in the case of no financing constraints. As discussed in Section 3.5.3, this effect is due to the flattening of the optimal coupon and may improve the predictive power of the existing structural models for credit spreads by addressing the critique of Eom et al. (2004).

The bottom two plots in Figure 3.5 display cash holdings versus leverage and credit spread for various levels of σ between 0.1 and 0.3. The relationships between cash and leverage (Figure 3.5.G), and cash and credit spread (Figure 3.5.H) is negative, which means that cash flow uncertainty may be the exogenous variable responsible for the similar regularities found in empirical studies (Opler et al. (1999), Acharya, Davydenko and Strebulaev (2007)).

3.5.5   Profitability uncertainty

Figure 3.6 reports the effects of changes in the uncertainty about the true level of a firm's expected EBIT, i.e., the second source of uncertainty in cash flows apart from σ. Under our assumption of the binomial distribution of $\overline{\mu}$, this uncertainty is captured by the spread between the high (μ_H) and low (μ_L) realizations of mean instantaneous earnings. This parameter captures the uncertain economic value of the firm (as opposed to σ, which describes the volatility of operational cash flows). In our comparative statics exercise, we vary $\mu_H - \mu_L$ around the mean $\mu_0 = \frac{1}{2}\left(\mu_H + \mu_L\right) = 0.1$. One effect is that a higher $\mu_H - \mu_L$ increases both the profit and the loss potential of the firm. The other effect is that with a higher spread $\mu_H - \mu_L$, the learning process μ_t becomes more rapid as the cash flow signals are more informative about either realization of $\overline{\mu}$. This leads to a decrease in default trigger μ^* (see Figure 3.2).

Figure 3.6.A reveals that the unconstrained firm increases the optimal coupon with rising $\mu_H - \mu_L$. This is the result of a lower default risk stemming from a decrease in μ^*. The rising c_u^* causes the debt value to increase in $\mu_H - \mu_L$ (Figure 3.6.B). It turns out that the equity value does not benefit from increased profit potential because this gain is offset by higher coupon payments (in Figure 3.6.C, the equity value of the unconstrained firm slowly decreases in $\mu_H - \mu_L$). The situation is different for the constrained firm. The driving force is the necessity to raise initial cash subject to the issuance costs. As before, the required cash is the lowest for intermediate coupons, so minimizing the flotation costs causes c^* to be driven to intermediate values (the solid line in 3.6.A). When c_u^* of the unconstrained firm is relatively low (here for low $\mu_H - \mu_L$), the financing constraints move the optimal coupon upwards. In such a case, the debt value increases and the equity value decreases. When c_u^* and c^* are already relatively high, the financing constraints prevent c^* from growing in $\mu_H - \mu_L$, which leads to a decrease in the debt value and an increase in the equity value. The effects on the debt and equity values are depicted in Figure 3.6.B and 3.6.C.

The forces just described for the case of the constrained firm are aggregated in Figure 3.6.D in the non-monotonic leverage ratio in $\mu_H - \mu_L$—first rising and then falling. The rising region is marked by low debt coupons, and the region of falling leverage is characterized by relatively high coupon payments. For the unconstrained firm, the profitability uncertainty $\mu_H - \mu_L$ has a positive effect on leverage, opposite to the effect of cash flow volatility σ.

Figure 3.6.E further shows that the cash holdings fall in increasing $\mu_H - \mu_L$. The negative effect comes from the increased speed of learning from cash flow shocks about the expected profitability. If negative liquidity translates quickly in a drop in μ_t, then less cash is required to cushion liquidity distress before insolvency at μ^*. It turns out
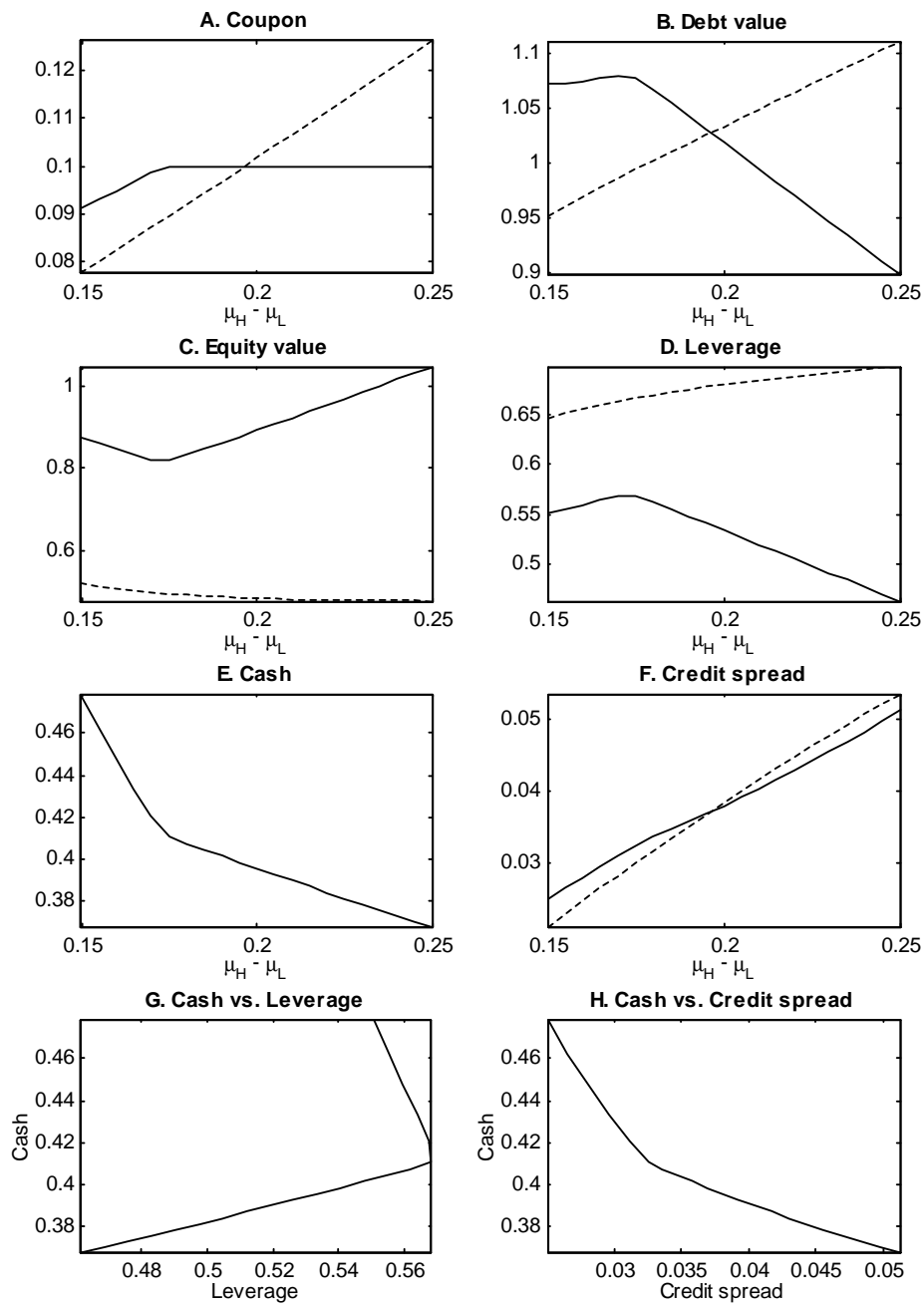
FIGURE 3.6. Effects of profitability uncertainty $\mu_H - \mu_L$ (mean preserving spread around $\mu_0 = 0.1$). The solid lines plot the respective values of the financially constrained firm with liquidity concerns, and the dashed lines plot the values of the unconstrained firm. Default and leverage are determined endogenously. The other parameter values are $\sigma = 0.2$, $r = 0.06$, $\tau = 0.15$, $\alpha = 0.6$ and $\lambda = 0.1$.

that this effect dominates, and $\overline{M}$ falls. The sign of the effect is again opposite to the effect of σ.

The analysis in this subsection, shown in Figure 3.6.F, reconfirms the findings from Figure 3.5.F, that the model with liquidity concerns and issuance cost predicts lower (higher) credit spreads when the model without financing constraints predicts relatively high (low) spreads.

In Figure 3.6.G and 3.6.H, cash holdings versus leverage and credit spread are shown for various levels of $\mu_H - \mu_L$ between 0.15 and 0.25. Notably, the effects just described lead to a negative relation between cash and credit risk. This means that the relationship preserves the same sign as in the case corresponding to σ uncertainty (cf. Figure 3.5). The relationship between cash and leverage is non-monotonic if the underlying changing exogenous variable is the uncertainty about profitability.

Our analysis draws attention to the need to differentiate between short-term volatility in cash flows and long-term uncertainty about economic prospects. It would be interesting to operationalize these measures of uncertainty and to test the predictions of our model empirically.

3.5.6   Leverage

A weakness of the standard trade-off model of capital structure that has frequently been raised in the literature is that the optimal leverage implied by the model exceeds the leverage ratios observed in empirical studies. Our model lessens this problem, as is best revealed in Figures 3.5 and 3.6. The second plots on the right-hand side in both figures present leverage ratios for various parameters. The leverage ratio of the firm with liquidity concerns is consistently seen to be significantly below the ratio of the unconstrained firm. While there are a number of effects that liquidity concerns bring to capital structure, the driving force behind this remarkably reduced leverage is the recognition of the role of cash in corporate assets. As the total assets of the constrained firm incorporate the value of cash, the leverage ratio decreases.

3.6   Conclusions

Earlier literature has studied either solvency default with optimal capital structure or liquidity default with cash and dividend policy separately. Our analytically tractable framework allows us to study a combination of both of these sources of financial distress and to enhance our understanding of the interaction of financing, cash, and dividends. With an extension to liquidity concerns, our setting addresses some of the weaknesses of the existing contingent claims models of corporate finance.

We believe that future research can use our model to study a number of additional issues. In order to stay reasonably focused we have concentrated on the analytically most tractable scenario under the assumption that a firm fully hedges its liquidity risk from the initial date. We leave it to future research to analyze the case of cash reserves below the target level $\overline{M}$. In this case, the firm may actually default because of either solvency or liquidity distress. We note here that such an analysis, by adding liquidity-driven default, could potentially alleviate the recognized problem of structural models based on solvency default—that is, the under-prediction of credit spreads and default probabilities for shorter horizons. It would also be interesting to extend our analysis to allow for different degrees of financing constraints. In such an extension, the firm would be able to raise new external financing beyond the initial date, but this financing would be subject to issuance costs. Finally, future research may incorporate corporate investments into the model to study the joint role of debt and cash in financing capital expansion in the presence of financing constraints.

3.A   Appendix: Proofs

Proof of Proposition 3.1. We first solve for the equity value function. Differential equation (3.5) has an analytical solution of the following general form:

$$E_u(\mu) = B_1 \left(\mu - \mu_L\right)^{1-\beta} \left(\mu_H - \mu\right)^{\beta} + B_2 \left(\mu - \mu_L\right)^{\beta} \left(\mu_H - \mu\right)^{1-\beta} + (1-\tau)\frac{\mu - c}{r}, \quad (3.32)$$

where $\beta > 1$ is the positive root of

$$\beta^2 - \beta - \frac{2r\sigma^2}{\left(\mu_H - \mu_L\right)^2} = 0,$$

and B_1, B_2 are constants that are determined by boundary conditions. (3.32) can be verified by direct substitution. The first two terms constitute the general solution to the homogenous part of (3.5) and the third term is an easy-to-guess particular solution to the whole non-homogenous equation (3.5). The boundary condition at μ_H implies that $B_2 = 0$. This is because, with $\beta > 1$ for any other B_2, $E_u(\mu_H)$ is unbounded. Using the boundary condition at μ_u^* to determine B_1 delivers the expression for $E_u(\mu)$ given in the proposition.

Debt value is found analogously using that the general solution to differential equation (3.4) is

$$D_u(\mu) = B_3 \left(\mu - \mu_L\right)^{1-\beta} \left(\mu_H - \mu\right)^{\beta} + B_4 \left(\mu - \mu_L\right)^{\beta} \left(\mu_H - \mu\right)^{1-\beta} + \frac{c}{r},$$

with β as above and constants B_3 and B_4. Applying the boundary conditions on D_u at μ_H and μ_u^* yields (3.10). Firm value F_u given in (3.11) follows by adding (3.9) and (3.10).

Optimal default trigger μ_u^* in (3.8) is delivered by applying the smooth pasting condition (3.7) to equation (3.9). ∎

Proof of Proposition 3.2. For an arbitrary function $\overline{M}(\cdot, \cdot)$, let $M_t = \overline{M}(\mu_t, X_t)$ so that M_t is allowed to depend on both state variables. Denote the default time associated with trigger μ^* by $t^* = \inf \{t \geq 0 : \mu_t < \mu^*\}$. The firm is liquid up to time t^* if $M_t \geq 0$ for all $t \leq t^*$. Note that, for example, a simple cash policy $M_t = 0$, $t \leq t^*$, satisfies this liquidity condition, but such a policy is not feasible as it requires negative dividends. From (3.14) we have

$$dDiv_t = rM_t dt - dM_t + dY_t. \tag{3.33}$$

The cash and dividend policy is feasible if the equality holds at each time. As the firm has full discretion over non-negative dividends, the cash policy remains feasible as long as $dDiv_t \geq 0$ in (3.33). We want to determine the lowest cash level $\overline{M}$ that satisfies both liquidity and feasibility conditions.

Suppose first that $\overline{M}(\mu, X)$ is a continuous and differentiable function. Applying Itô's lemma to $\overline{M}$, the right-hand side of (3.33) can be written as

$$
\begin{aligned}
dDiv_t = & \left[r\overline{M} + (1 - \tau)(\mu_t - c) - \frac{1}{2\sigma^2}(\mu_t - \mu_L)^2 (\mu_H - \mu_t)^2 \overline{M}_{\mu\mu} - \mu \overline{M}_X \right. \\
& \left. - \frac{1}{2}\sigma^2 \overline{M}_{XX} - (\mu_t - \mu_L)(\mu_H - \mu_t)\overline{M}_{\mu X} \right] dt \\
& + \left[(1 - \tau)\sigma - \frac{1}{\sigma}(\mu_t - \mu_L)(\mu_H - \mu_t)\overline{M}_\mu - \sigma \overline{M}_X \right] dZ_t, \quad (3.34)
\end{aligned}
$$

where subindexes at $\overline{M}$ denote partial derivatives. Our requirement that increments of this process are non-negative for all $t \leq t^*$ can be satisfied if and only if, first, the volatility coefficient at dZ_t is constant and zero and, second, the drift parameter at dt is non-negative. The first condition yields the following partial differential equation:

$$\frac{1}{\sigma^2}(\mu_t - \mu_L)(\mu_H - \mu_t)\overline{M}_\mu + \overline{M}_X = (1 - \tau). \tag{3.35}$$

Its general solution is

$$\overline{M}(\mu, X) = (1 - \tau)\frac{\sigma^2}{\mu_H - \mu_L}\ln\left(\frac{\mu - \mu_L}{\mu_H - \mu}\right) + C_1\left[\frac{\mu_H - \mu_L}{\sigma^2}X - \ln\left(\frac{\mu - \mu_L}{\mu_H - \mu}\right)\right] + C_2, \tag{3.36}$$

where C_1 and C_2 are constants. As X_t, $t \leq t^*$, can in general take any positive or negative values, the liquidity condition $M_t \geq 0$, $t \leq t^*$, is satisfied only if $C_1 = 0$. This means that $\overline{M}$ is independent of X. To determine C_2, we use the non-negativity condition on the drift parameter in (3.34), which, with the use of (3.36), can be written as

$$r\overline{M}(\mu, X) + (1 - \tau)\left(\frac{\mu_H + \mu_L}{2} - c\right) \geq 0.$$

We note that $\overline{M}$ is increasing in μ, which implies that the inequality is most demanding at $\mu = \mu^*$. Moreover, the liquidity condition at all $t \leq t^*$ requires that

$$\overline{M}(\mu^*, X) \geq 0.$$

Solving the last two inequalities for the constant C_2, we obtain the formula given in the proposition.

Finally, we rule out that there are points of discontinuity and non-differentiability in $\overline{M}$ if $\mu > \mu^*$. If $\overline{M}$ is discontinuous, it can only have downward jumps. But if immediately after the jump $\overline{M}$ is the smallest M that allows the firm to avoid liquidity default, then in the continuous environment of the model, $\overline{M}$ before the jump could not be the smallest M satisfying this desired property. Hence $\overline{M}$ must be continuous. Suppose now that $\overline{M}$ has some non-differentiable points. In between the points, $\overline{M}$ must satisfy differential equation (3.35) with the general solution in (3.36), subject to the boundary conditions implied by the continuity of $\overline{M}$. But with $C_1 = 0$, it will result in $\overline{M}$ that is a continuous differentiable function of μ for all $\mu > \mu^*$. ∎

Proof of Proposition 3.3. For a given μ^*, define the value function $E(\mu, M)$ as follows. For $\mu > \mu^*$ and $0 < M < \overline{M}(\mu)$, $E(\mu, M)$ satisfies the differential equation

$$rE(\mu, M) = \frac{1}{2\sigma^2}(\mu_t - \mu_L)^2(\mu_H - \mu_t)^2 E_{\mu\mu}(\mu, M) + \frac{1}{2}(1-\tau)^2\sigma^2 E_{MM}(\mu, M)$$
$$+ (1-\tau)(\mu - \mu_L)(\mu_H - \mu_t)E_{\mu M}(\mu, M) + [rM + (1-\tau)(\mu - c)]E_M(\mu, M), \quad (3.37)$$

For $\mu \geq \mu^*$ and $M \geq \overline{M}(\mu)$, $E(\mu, M)$ is given by

$$E(\mu, M) = E_u(\mu) + M. \tag{3.38}$$

For if $\mu \geq \mu^*$ and $M = 0$

$$E(\mu, 0) = 0.$$

To prove that the policy specified in (3.19)-(3.21) attains $E(\mu, M)$ and that no other feasible policy provides a higher value, define for any feasible payout policy Div_t

$$W_t = \int_0^t e^{-rs} dDiv_s + e^{-rt} E(\mu_t, M_t).$$

If payout Div_t is continuous at t, then, from Itô's lemma applied to $E_t = E(\mu_t, M_t)$,

$$
\begin{aligned}
dW_t &= e^{-rt} dDiv_t - re^{-rt} E(\mu_t, M_t) dt + e^{-rt} dE_t \\
&= e^{-rt} dDiv_t - re^{-rt} E(\mu_t, M_t) dt + e^{-rt} \left\{ \frac{1}{2\sigma^2} (\mu_t - \mu_L)^2 (\mu_H - \mu_t)^2 E_{\mu\mu}(\mu_t, M_t) \right. \\
&\quad + \frac{1}{2} (1 - \tau)^2 \sigma^2 E_{MM}(\mu_t, M_t) + (1 - \tau)(\mu - \mu_L)(\mu_H - \mu_t) E_{\mu M}(\mu_t, M_t) \\
&\quad \left. + [rM + (1 - \tau)(\mu - c)] E_M(\mu_t, M_t) \right\} dt - e^{-rt} dDiv_t E_M(\mu_t, M_t) \\
&\quad + e^{-rt} \left[\frac{1}{\sigma}(\mu_t - \mu_L)(\mu_H - \mu_t) E_\mu(\mu_t, M_t) + (1 - \tau)\sigma E_\mu(\mu_t, M_t) \right] dZ_t \\
&= e^{-rt} \left(1 - E_M(\mu_t, M_t) \right) dDiv_t \\
&\quad + e^{-rt} \left[\frac{1}{\sigma}(\mu_t - \mu_L)(\mu_H - \mu_t) E_\mu(\mu_t, M_t) + (1 - \tau)\sigma E_\mu(\mu_t, M_t) \right] dZ_t, \qquad (3.39)
\end{aligned}
$$

where for the last equality we use (3.37) if $0 < M < \overline{M}(\mu)$ or (3.38) combined with (3.5) if $M \geq \overline{M}(\mu)$.

Note that W_t is a martingale if $(1 - E_M(\mu_t, M_t)) dDiv_t$ equals 0. As $E_M(\mu_t, M_t) \geq 1$ if $M_t < \overline{M}(\mu_t)$ (by (3.22)) and $E_M(\mu_t, M_t) = 1$ if $M_t \geq \overline{M}(\mu_t)$ (by 3.38), the policy proposed in (3.19)-(3.21) guarantees that W_t is a martingale. This implies that the value that is obtained by the equity holders from the dividend distribution specified in (3.19)-(3.21) is equal to $E(\mu_0, M_0)$. Indeed,

$$\mathbb{E} \left[\int_0^{t^*} e^{-rs} dDiv_s^* \right] = \mathbb{E}[W_{t^*}] = W_0 = E(\mu_0, M_0),$$

where the second equality holds because W_t is a martingale.

For any other feasible payout policy it must hold that $dDiv_t \geq 0$ and $E_M(\mu_t, M_t) \geq 1$. It follows that the drift of W_t, $(1 - E_M(\mu_t, M_t)) dDiv_t$, is non-positive and thus W_t is a supermartingale. Consequently,

$$\mathbb{E} \left[\int_0^{t^*} e^{-rs} dDiv_s \right] = \mathbb{E}[W_{t^*}] \leq W_0 = E(\mu_0, M_0), \qquad (3.40)$$

so the present value of dividend payouts in this alternative policy is less than or at most equal to $E(\mu_0, M_0)$.

There are no jumps in Div_t^* for $t \in (0, t^*)$, so the argument above is complete with respect to Div_t^*. If there is a jump of $\Delta Div_t > 0$ in an alternative payout Div_t for some $t \in (0, t^*)$, then (3.39) does not apply at t. But then W_t jumps by $e^{-rt}\Delta Div_t + e^{-rt}E(\mu_t, M_t - \Delta Div_t) \leq 0$ as $E_M(\mu_t, M_t) \geq 1$. Thus W_t is a supermartingale and the argument in (3.40) applies. ∎

Proof of Proposition 3.4. Debt value is found as in the proof of Proposition 3.1. To determine equity value, we use the general solution to differential equation (3.23). By direct verification, it is

$$E(\mu) = B_5 \left(\mu - \mu_L\right)^{1-\beta} \left(\mu_H - \mu\right)^{\beta} + B_6 \left(\mu - \mu_L\right)^{\beta} \left(\mu_H - \mu\right)^{1-\beta}$$
$$+ \frac{a_1}{r} + \frac{a_2}{r} \left[\ln \frac{\mu - \mu_L}{\mu_H - \mu} + \frac{\mu_H - \mu_L}{r\sigma^2} \left(\mu - \frac{\mu_H - \mu_L}{2}\right)\right].$$

Applying the boundary conditions at μ_H and μ^* to determine constants B_5 and B_6, we obtain the expression provided in the proposition. ∎

4

Finite Project Life and Uncertainty Effects on Investment

4.1 Introduction

The standard theory of the real options approach to investment, as explained and reviewed in Dixit and Pindyck (1994)[1], states that uncertainty in combination with irreversibility creates a value of the option to wait with undertaking capital investments. Over time more information becomes available, which enables the decision maker to make better investment decisions at a later date.

The general prediction of the real options literature is that a higher level of uncertainty increases the value of waiting and thus has a negative effect on investment. In this chapter we revisit this conclusion. To do so we adopt the standard framework with contingent claims valuation of the investment opportunity and change it in one aspect: where the vast majority of the real options literature assumes projects to be perpetual, we allow for the project to generate earnings only during a finite amount of time.[2] The assumption of a project having an infinite life is useful mostly due to its simplicity. However, in corporate practice the investment projects are usually considered to have a finite life. This is especially true in today's knowledge economy, in which innovations limit the economic lifetime of technologies.[3] We show that the simplifying assump-

---

[1] Some more recent contributions include studies of implications of learning (Décamps and Mariotti (2004), Thijssen, Huisman and Kort (2006)), agency (Grenadier and Wang (2005)), strategic quality choice (Pennings (2004)), business cycle (Guo, Miao and Morellec (2005)), policy change (Pawlina and Kort (2005)), and implications to capital structure choices (Miao (2005)), mergers and acquisitions dynamics (Morellec and Zhdanov (2005)), or exit strategies (Murto (2004)).

[2] Notably, Majd and Pindyck (1987) discuss some implications of finite project life on real options modeling. While they provide some arguments and cases when the finite project life considerations can be omitted, these considerations turn out to play an important role in our analysis.

[3] Certainly, the same arguments point toward introducing a finite life of the investment opportunity and not only of the project after investment. We do study this case in Section 4.4.1 where it is shown that our main result also holds there.

tion of perpetual projects is critical for the investment-uncertainty relationship. Our main result is that the investment threshold decreases with uncertainty in case the uncertainty level is low and the project life is short. So, changing the project life from infinite to finite can imply a negative relationship between uncertainty and the value of waiting, which reverses the basic real options result.

To be more precise, an increase in uncertainty affects the investment threshold in three different ways. The first effect is the *discounting effect*. An increase of uncertainty raises the discount rate via the risk premium component. This reduces the net present value of the investment and thus raises the investment threshold. The second effect is the *volatility effect*, which affects the value of the option to wait positively: higher uncertainty increases the upside potential payoff from the option, leaving the downside payoff unchanged at zero (since the option will not be exercised at low payoff values). This increased option value implies that the firm has more incentive to wait, which also increases the investment threshold. The third effect of an increase of uncertainty on the investment threshold is the *convenience yield effect*. The increase of asset riskiness raises the discount rate and thus also the convenience yield of the investment opportunity. This decreases the value of waiting, so that it is more attractive to invest earlier resulting in a lower investment threshold.

The discounting and volatility effects thus raise the investment threshold, while the convenience yield effect works in the opposite direction. Projects with a short life are relatively insensitive to discount rates. On the other hand, at low levels of uncertainty, increased uncertainty has still little effect on the probability of observing low prices, and thus the volatility effect is small in this case. Consequently, it is possible for the negative convenience yield effect to dominate the two other effects when the project life is finite and uncertainty is low. In that case it thus holds that the investment threshold decreases with uncertainty.

We examine the robustness of the non-monotonic effect of uncertainty on investment in the case of a finite project life by considering several variations of the problem. First, we show that this result survives in case the opportunity to invest in the project is available only for a limited amount of time. Next, we prove that this also holds for other relaxations of the infinite project life assumption, like uncertain project duration and capital depreciation. Furthermore, we find that generalized functional forms of the convenience yield preserve the observed relationships. Finally, the non-monotonic effect is also present in case revenues are mean reverting.

The impact of uncertainty on investments has been of interest to economists for a long time. One strand of literature relies on convex costs of capital adjustment and convexity of marginal profits in prices. As shown by Hartman (1972) and Abel (1983), in such a setting uncertainty hastens investment. The other important strand

of literature, based on the real options theory, acknowledges (partial) irreversibility of investments and predicts that uncertainty delays investment. This chapter verifies the latter prediction and shows that the investment trigger is not necessarily increasing in uncertainty. Most closely related papers are Caballero (1991) and Bar-Ilan and Strange (1996). Caballero (1991) considers a perfect competition setting with convex adjustment costs, and he obtains that irreversibility does not lead to the usual negative investment-uncertainty relationship. Bar-Ilan and Strange (1996) assume that there are lags between investment decisions and realizations. Firms have abilities to abandon uncompleted projects in bad times, which creates a convexity in the output and value functions. Bar-Ilan and Strange (1996) find that uncertainty may accelerate as well as decelerate investment depending on specific parameter values. Both papers have in common that they depart from the conventional result of the real options literature, because the models create convexities in line of Hartman (1972) and Abel (1983). Thus it comes with little surprise that in these papers uncertainty may either accelerate or decelerate investment. The result of our analysis is unique in the sense that uncertainty may hasten irreversible investment without building on the convexity of the marginal product of capital. Our model remains in the pure real options framework and the reversal of the conventional result builds solely on the contingent claims valuation of investment opportunities and a finite capital lifetime. Moreover, since we only depart from the standard real option framework by imposing a finite lifetime, our model is more general and is thus applicable to more investment situations than the models in Caballero (1991) and Bar-Ilan and Strange (1996).

A different approach to study the relationship between uncertainty and irreversible investments is taken by Sarkar (2000). Sarkar analyzes the probability of investment taking place within a certain time period and points at the fact that an increasing trigger does not automatically mean that investment will be delayed. The difference with our result is that we show that increased uncertainty may not even lead to an increased trigger.

Beyond this introduction the chapter is organized as follows. In the next section we consider the model of the finitely-lived project and derive the optimal investment trigger. Section 4.3 studies how uncertainty influences the investment decision. In Section 4.4 we discuss robustness, while Section 4.5 concludes. All proofs are contained in Appendix 4.A.

4.2   The model and the optimal investment decision

We consider an irreversible investment project with finite life of T years that can be undertaken at any time. After the investment has taken place, the project generates a

stochastic revenue of Q_t per unit time. Q_t evolves exogenously according to a geometric Brownian motion

$$dQ_t = \mu Q_t dt + \sigma Q_t dZ_t, \tag{4.1}$$

where dZ is the increment of a standard Wiener process, μ is the drift parameter and σ is the volatility parameter that introduces the uncertainty in our model. Throughout the chapter we assume that $\mu, \sigma > 0$. When the project is undertaken, a one-time investment cost I is paid. For simplicity, the marginal costs are put equal to zero.

We employ the contingent claims approach to real options valuation.[4] Under the standard assumption of market completeness, the expected rate of return of the project π is determined in the financial market equilibrium. The CAPM[5] formula relates π, the risk-free interest rate r, the correlation of the project return with the return of the market portfolio ρ, and the market price of risk λ as follows:

$$\pi = r + \lambda \rho \sigma. \tag{4.2}$$

The difference between π, the expected return of the project, and μ, the expected rate of change of Q, is referred to as the convenience yield (or return shortfall) of the investment opportunity. The later is denoted by δ and satisfies

$$\delta \equiv \pi - \mu = r + \lambda \rho \sigma - \mu. \tag{4.3}$$

We assume that $\delta > 0$, which ensures that the investment is ever undertaken; otherwise it is never optimal to exercise the option.

The level of uncertainty faced by the firm is measured by the volatility parameter σ. From (4.3) we obtain that a change in σ results in a change of π, which must lead to an adjustment of either μ or δ or both. In general, this relation depends on what is assumed to be an endogenous parameter affected by changes in volatility. A certain guideline in this respect could be Pindyck (2004), that relates commodity inventories, spot and future prices and the level of volatility. The model is estimated for several commodities and the results show that a volatility shock has a significant effect on the convenience yield and only a small effect on the price. Consistent with this evidence, it also seems to be more common in the related literature on the investment-uncertainty

---

[4] The standard methods in real options theory to value an investment opportunity are dynamic programming and contingent claims valuation (Dixit and Pindyck (1994)). Compared to dynamic programming, the contingent claims approach offers a better treatment of the discount rate, because it is endogenously determined as an implication of the overall equilibrium in capital markets.

[5] The assumption that the intertemporal Capital Asset Pricing Model (CAPM) of Merton (1973) holds is in accordance with the related literature. The CAPM brings a linear relationship between the discount rate and σ. In Section 4.4.3 we show that the results we present also hold for more general discount rate functions.

relationship to assume that μ is fixed and δ changes with σ (e.g. Sarkar (2000) and Sarkar (2003)). We follow this modeling convention.

The value of the project $V(Q)$ evolves over time and depends on the current realization of Q. Upon installation the project value is equal to the expected present value of the revenue stream discounted by the risk-adjusted discount rate. If the project has a finite life of T years, then the project value at the time of the investment is

$$V(Q) = \mathbb{E}\left[\int_0^T e^{-\pi t} Q_t dt \,|\, Q_0 = Q\right] = \int_0^T e^{-(\pi - \mu)t} Q dt = Q \frac{1 - e^{-(r + \lambda\rho\sigma - \mu)T}}{r + \lambda\rho\sigma - \mu}. \quad (4.4)$$

Before the project is installed, the firm holds an option to invest. The option is held until the stochastic revenue flow reaches a sufficiently high level at which it is optimal to exercise the option and invest. The option value $F(Q)$ can be found by the replicating portfolio argument. Employing the standard methods (cf. Dixit and Pindyck (1994)) yields that $F(Q)$ must satisfy the differential equation:

$$\frac{1}{2}\sigma^2 Q^2 F''(Q) + (\mu - \lambda\rho\sigma)Q F'(Q) - rF(Q) = 0. \quad (4.5)$$

We solve this differential equation subject to the value matching and smooth pasting conditions at the investment trigger Q^* and a zero value condition at $Q = 0$. The derivations are standard and are omitted here. The resulting firm value prior to investment is

$$F(Q) = (V(Q^*) - I)\left(\frac{Q}{Q^*}\right)^{\beta_1}.$$

The optimal investment rule is given by the upper trigger

$$Q^* = \frac{\beta_1}{\beta_1 - 1} \frac{r + \lambda\rho\sigma - \mu}{1 - e^{-(r + \lambda\rho\sigma - \mu)T}} I, \quad (4.6)$$

while β_1 is the positive root of the quadratic equation

$$L_0 \equiv \frac{1}{2}\sigma^2 \beta(\beta - 1) + (\mu - \lambda\rho\sigma)\beta - r = 0, \quad (4.7)$$

and equals

$$\beta_1 = \frac{1}{2} - \frac{\mu - \lambda\rho\sigma}{\sigma^2} + \sqrt{\left(\frac{\mu - \lambda\rho\sigma}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}}. \quad (4.8)$$

Under the net present value (NPV) rule the investment is undertaken as soon as the risk-adjusted project value exceeds the investment cost, that is at the revenue level equal to $\frac{r + \lambda\rho\sigma - \mu}{1 - e^{-(r + \lambda\rho\sigma - \mu)T}} I$. This value is always lower than Q^* in (4.6), as $\beta_1 > 1$. So there

are states where the expected payoff of investment is positive and the firm chooses to wait and not to invest. The option to invest captures this positive value of waiting.

4.3   The effects of uncertainty on the investment trigger

This section studies the effect of uncertainty on the value of waiting. First, we show that, as usual, the value of waiting, reflected in the level of investment trigger, always increases with uncertainty when the project life is infinite or when discount rates are unaffected by uncertainty. Second, if the equilibrium discount rate contains a positive risk premium, we derive that the value of waiting decreases with uncertainty in case of finite project lives and low levels of uncertainty. Finally, we provide an economic analysis of these results.

4.3.1   Monotonicity results

We start out with the basic real options result for the investment project with infinite life.

Proposition 4.1 *If the project life is infinite, the investment trigger increases with uncertainty.*

In case of an infinite project life the effect of uncertainty on the investment trigger is unambiguously positive. This is the standard real options result, which says that the value of waiting increases with uncertainty. This is reflected by higher trigger values, because the revenue must reach a higher level before investment is optimally undertaken.

Now, let us move on to the finite life project case. We first consider the scenario where the impact of systematic risk is absent or not priced by the market. This implies that the discount rate is constant, and requires that either the market price of risk is zero, $\lambda = 0$, or that the correlation of the project return with the return of the market portfolio is zero, $\rho = 0$.

Proposition 4.2 *If the discount rate is constant (zero market price of risk or zero correlation of project return with the return of the market portfolio), the relationship between the investment trigger and uncertainty is always positive.*

Proposition 4.2 states that, in the absence of the risk premium effect the investment trigger always increases with uncertainty irrespective of the project lifetime, which is again the usual real options result. It is important to point out, however, that the conditions necessary for constant discount rates are in general difficult to accept in the context of investment models; see discussions in e.g. Zeira (1990) and Sarkar (2003).

The next proposition considers one case where the discount rate is not constant.

Proposition 4.3 *If $\lambda\rho < 0$, then the relationship between the investment trigger and uncertainty is always positive.*

This result shows that in case of a negative risk premium (possible if either the correlation of the project return with the return of the market portfolio or the market price of risk is negative), the usual positive relationship arises.

4.3.2   Non-monotonicity result

We proved in the previous subsection that both in the model with a project of infinite life and in the model without a risk premium or with a negative risk premium, the impact of uncertainty on the investment trigger is always positive. These are interesting special and limit cases; however, the assumptions of Propositions 4.1 and 4.2 are serious abstractions from reality, and the negative risk premium condition of Proposition 4.3 is a relatively rare phenomenon in the markets. Next, we turn to the most common situation where the project life is finite and the discount rate is set in the capital market equilibrium with a positive risk premium. We now show that the effect on the trigger is no longer monotonic in uncertainty.

Proposition 4.4 *If the project life is finite and $\lambda\rho > 0$, the uncertainty effect on the investment trigger is non-monotonic: it decreases in σ for low levels of σ and then increases. The length of the σ-interval where the negative effect occurs is negatively related to the project lifetime.*

Figure 4.1 presents some numerical examples, where the parameter values correspond to earlier work on the investment-uncertainty relationship, in particular to Sarkar (2000). We see that indeed there is a negative relation between σ and Q^* for lower values of σ. The effect is more pronounced for short-term projects, but even in the case of a 30-year project Q^* decreases until σ is around 0.12. The example shows that the positive effect of uncertainty on investment (negative on the trigger) arises for economically relevant parameter values. The figure, of course, also confirms that for an infinitely long project the relation is monotonic and increasing in line of the results in Proposition 4.1.

4.3.3   Economic analysis of the non-monotonicity result

In this section we provide an economic interpretation of the non-monotonic effect of uncertainty shown in Proposition 4.4 (we assume here that $\lambda\rho > 0$). From (4.3) and
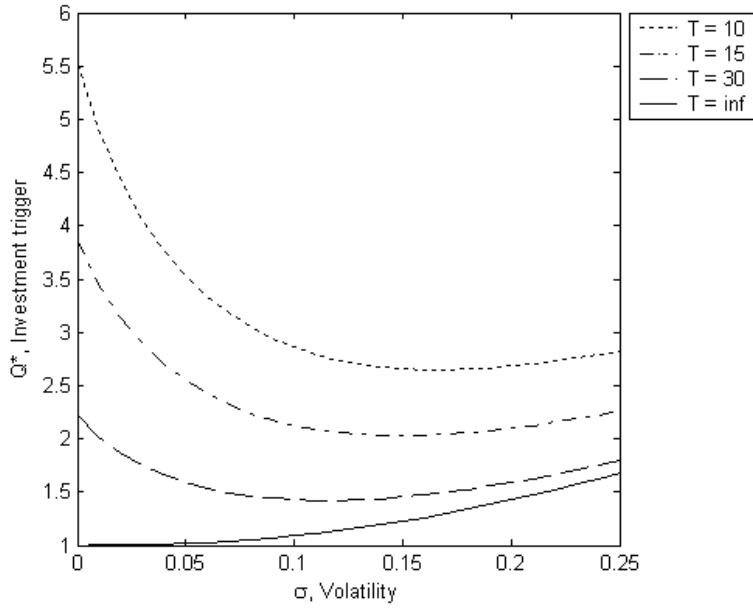
FIGURE 4.1. Investment trigger as a function of volatility for various project lenghts T and the set of parameters: $\mu = 0.08$, $r = 0.1$, $\rho = 0.7$, $\lambda = 0.4$, $I = 10$.

(4.6) it follows that the investment trigger can also be expressed as

$$Q^* = \frac{\beta_1}{\beta_1 - 1} \frac{\delta}{1 - e^{-\delta T}} I. \tag{4.9}$$

At this point it is convenient to trace all the variables that are affected by uncertainty and consider the trigger as a function of three parameters: $Q^*(\sigma, \delta(\sigma), \beta_1(\sigma, \delta(\sigma)))$. Then the derivative of the investment trigger with respect to σ can be decomposed into three effects in the following way:

$$\frac{d}{d\sigma} Q^*(\sigma, \delta(\sigma), \beta_1(\sigma, \delta(\sigma))) = \underbrace{\frac{\partial Q^*}{\partial \delta} \frac{\partial \delta}{\partial \sigma}}_{\text{Discounting effect}} + \underbrace{\frac{\partial Q^*}{\partial \beta_1} \frac{\partial \beta_1}{\partial \sigma}}_{\text{Volatility effect}} + \underbrace{\frac{\partial Q^*}{\partial \beta_1} \frac{\partial \beta_1}{\partial \delta} \frac{\partial \delta}{\partial \sigma}}_{\text{Convenience yield effect}}. \tag{4.10}$$

The three effects have a clear interpretation and each has an unambiguous sign (for the case of $\lambda\rho > 0$). The first term on the right-hand side, the *discounting effect*, measures the impact of revenue uncertainty on the rate used to discount the project value. Rising uncertainty increases the discount rate, which reduces the net present value of the investment project. This implies that it is less profitable to invest in this project, which leads to an increase of the trigger value. Consequently, as is straightforward to derive, the discounting effect is always positive.

Since the second and the third term of (4.10) both affect the trigger value via β_1, they reflect the influence of uncertainty on the value of the option to wait. Below we

refer to these two effects combined as the *option effect*. The *volatility effect*, which is represented by the derivative $\frac{\partial Q^*}{\partial \beta_1} \frac{\partial \beta_1}{\partial \sigma}$, captures the direct impact of uncertainty on the value of the option to wait. Higher uncertainty increases the upside potential payoff from the option, leaving the downside payoff unchanged at zero (since the option will not be exercised at low payoff values). This is the well-known positive impact of uncertainty on the option value. An increased option value implies that the firm has more incentive to wait. This raises the opportunity cost of investing so that the investment trigger will increase. Hence, the effect is unequivocally positive.

The product $\frac{\partial Q^*}{\partial \beta_1} \frac{\partial \beta_1}{\partial \delta} \frac{\partial \delta}{\partial \sigma}$ in (4.10) represents the impact of uncertainty on the option value through the convenience yield, and we refer to it as the *convenience yield effect*. Increased uncertainty raises the risk premium of the expected rate of return and thus also the convenience yield, which in turn elevates the opportunity cost of holding the option and consequently decreases its value. For this reason it is attractive to invest earlier, which reduces the trigger.

Summarizing, we conclude that the discounting and volatility effects are positive, while the convenience yield effect is negative. Clearly, the negative relationship between uncertainty and investment occurs only if the convenience effect dominates the two other effects. The following proposition shows how the uncertainty level and the project length influence the relative size of the three effects.

Proposition 4.5 *(i) Define $\hat{\sigma} = \{\sigma \geq 0 : (\beta_1 - 1)\sigma - \lambda\rho = 0\}$. The combined option effect is negative at $\sigma < \hat{\sigma}$ and positive at $\sigma > \hat{\sigma}$.*
(ii) The shorter is the project life T, the smaller is the discounting effect and the larger in absolute terms are the two option effects.

The first part of the proposition states that the sign of the effect of uncertainty on the option value is ambiguous but separable into two regions.[6] At a relatively high uncertainty level the positive volatility effect dominates the negative convenience yield effect. At low levels of uncertainty the negative effect dominates. In such a case, a marginal increase in uncertainty has little impact on the probability of reaching extreme values by the underlying process and hence the volatility effect is relatively small. On the other hand, the convenience yield effect is also significantly present at low levels of uncertainty, since the convenience yield δ is linear in σ, implying that the marginal effect of σ in δ is constant (in fact the convenience yield effect is not constant but diminishes at higher σ, as the full effect works via the discount factor).

---

[6]From Proposition 4.5 (i) it is clear that in a setup where only the option effects are present, the non-monotonic investment-uncertainty relationship would arise irrespective of the project lifetime. This could be the case for example, if the project value V behaves according to geometric Brownian process. This was shown in a contemporaneous work by Wong (2007). However such a setup is a rather serious abstraction from reality (see Dixit and Pindyck (1994; p. 175) for arguments) and the negative effect disappears as soon as perpetual revenues from the project are directly modelled.

The second part of the proposition states that the project and option-related effects react differently to changes in the project life. The discounting effect becomes smaller with shorter project lives. Clearly, short-lived projects are relatively insensitive to marginal changes of the discount rate. On the other hand, the option-related effects increase with shorter project lives. This is because a shorter project life implies that the current revenue flow needs to be larger for the investment to be optimal, which leads to larger option effects.

Now we are ready to establish when and why an increasing uncertainty level may lower the investment threshold. At low levels of uncertainty, the positive volatility effect is small and the effects working via discount rate and convenience yield are still significant. These two last effects have opposing signs so that a low σ alone is not enough to observe a negative total effect (cf. Proposition 4.1). If, however, in addition the project life is short then the positive discounting effect will be small and the negative convenience yield effect dominates. Therefore, at low levels of σ and T, it is possible that the negative convenience yield effect dominates the two positive effects (see Proposition 4.4).

These mechanisms are illustrated in a numerical example presented in Table 4.1. It allows for a closer inspection of the magnitude of the effects of uncertainty affecting the position of the investment trigger. The volatility and convenience yield effects increase with shortening the project life. The discounting effect decreases with smaller T. The combined option effect is negative for low levels of σ but it is increasing in σ (it becomes positive for $\sigma > \hat{\sigma} = 0.241$). The longer the project life, the faster is the negative convenience yield effect offset by the positive impact of the discounting and volatility effects. If $T = 10$, the total effect is negative for σ between 0 and 0.16, while for $T = 30$ the total effect remains negative for σ between 0 and 0.10.

4.4 Robustness

The model of the previous sections has been geared to show our results in the simplest setting. The aim of this section is to demonstrate that our main result, i.e. that the value of waiting decreases with uncertainty in case of a short project life and a limited amount of uncertainty, can be generalized. First we consider a scenario where the investment opportunity is available only for a limited amount of time. After that we analyze the case where the project has an uncertain duration. Next, we consider more general, thus not necessary linear, convenience yield functions in uncertainty. Finally, we allow the revenue process to be mean reverting.

σ	$T = 10$					$T = 30$				
	Q^*	(1)	(2)	(3)	(4)	Q^*	(1)	(2)	(3)	(4)
0.00	5.52	7.47	0.00	-77.23	-69.77	2.22	8.38	0.00	-31.03	-22.65
0.02	4.44	5.95	1.19	-48.51	-41.37	1.87	6.86	0.50	-20.44	-13.08
0.04	3.77	5.01	2.15	-33.70	-26.54	1.66	5.92	0.95	-14.86	-8.00
0.06	3.34	4.39	2.99	-25.03	-17.65	1.54	5.29	1.38	-11.53	-4.86
0.08	3.05	3.97	3.75	-19.47	-11.75	1.46	4.88	1.80	-9.35	-2.67
0.10	2.86	3.68	4.40	-15.66	-7.57	1.43	4.60	2.20	-7.82	-1.02
0.12	2.74	3.49	4.95	-12.92	-4.48	1.42	4.43	2.57	-6.70	0.29
0.14	2.67	3.38	5.36	-10.90	-2.17	1.44	4.34	2.88	-5.86	1.36
0.16	2.65	3.31	5.65	-9.38	-0.42	1.47	4.30	3.15	-5.22	2.22
0.18	2.65	3.28	5.85	-8.24	0.90	1.53	4.30	3.37	-4.74	2.93
0.20	2.68	3.28	5.99	-7.37	1.91	1.59	4.34	3.55	-4.37	3.52
0.22	2.73	3.31	6.09	-6.70	2.69	1.67	4.39	3.72	-4.09	4.02
0.24	2.79	3.34	6.17	-6.19	3.32	1.75	4.46	3.87	-3.89	4.45

TABLE 4.1. The three effects of uncertainty affecting the position of the investment trigger for the set of parameters: $\mu = 0.08$, $r = 0.1$, $\rho = 0.7$, $\lambda = 0.4$, $I = 10$, $Q = 1$. The columns present: the discounting effect (1), the volatility effect (2), the convenience yield effect (3), and the total effect (4).

4.4.1 Finite-life option

We now assume that the project and the option to invest both have finite durability. McDonald and Siegel (1986) also allow for a finite life of the investment opportunity, but their project is implicitly perpetual. Finite life options have been extensively studied, and the book by Detemple (2005) provides background on recent analytical, approximation and numerical methods.

The project life is T years and its value $V(Q)$ is given by equation (4.4). Denote the life length of the option as T_F. Since the option expires at T_F, its value $F(Q, \tau_F)$ depends on remaining time τ_F to maturity. To find the differential equation defining the option value we follow the same steps as in Section 4.2. The resulting partial differential equation includes the time derivative and is given by

$$\frac{1}{2}\sigma^2 Q^2 F_{QQ} + (\mu - \lambda\rho\sigma)Q F_Q - F_\tau - rF = 0. \qquad (4.11)$$

The option value must satisfy the terminal condition at the expiry date T_F :

$$F(Q, 0) = \max\left(V(Q) - I, 0\right),$$

which states that at $\tau = 0$ the option is exercised (the investment is undertaken) if the project's expected present value exceeds the investment cost. The option satisfies also the boundary conditions at $Q = 0$ and $Q = Q^*$ similar to the ones used in Section 4.2: $F(Q^*, \tau_F) = V(Q^*) - I$, $F_Q(Q^*, \tau_F) = V'(Q^*)$ and $F(0, \tau_F) = 0$.

FIGURE 4.2. Project and option with finite life: investment trigger boundary, $Q^*(\tau_F)$, for various levels of volatility and the set of parameters: $\mu = 0.08$, $r = 0.1$, $\rho = 0.7$, $\lambda = 0.4$, $I = 10$, $T = 10$, $T_F = 10$.

Unlike in the previous problem, in which Q^* was a single point, here the optimal investment trigger $Q^*(\tau_F)$ is a function of time. The problem we have to solve is analogous to the valuation of American-style options with a finite expiry date, to which no closed-form solutions exist. We numerically solve equation (4.11) together with the boundary conditions using the Crank-Nicholson finite-difference scheme. We apply the logarithmic change of variable and use a mesh size of 500×500 points.

Figures 4.2 and 4.3 present our results for the optimal investment trigger boundary $Q^*(\tau_F)$. We assumed the option life T_F to be 10 years and the project life T to be either 10 years (Figure 4.2), or perpetual (Figure 4.3). All other parameters are as in the numerical example of Figure 4.1. The triggers $Q^*(\tau_F)$ are drawn for various levels of σ ranging from 0.10 to 0.30. The horizontal axis depicts the remaining option life τ_F.

As expected, the right-hand-side of both figures at $\tau_F = T_F = 10$ is well approximated by the model with a perpetual real option, so that the trigger boundary values are very close to those in Figure 4.1 ($T = 10$ and $T = \infty$ curves). At $\tau_F = 0$, when the investment decision becomes a now-or-never decision, all curves are at the values implied by the NPV investment rule.

Figure 4.2 clearly confirms our result that a finite project life may cause the real option investment rule to be non-monotonic in uncertainty. An increase of σ from 0.10

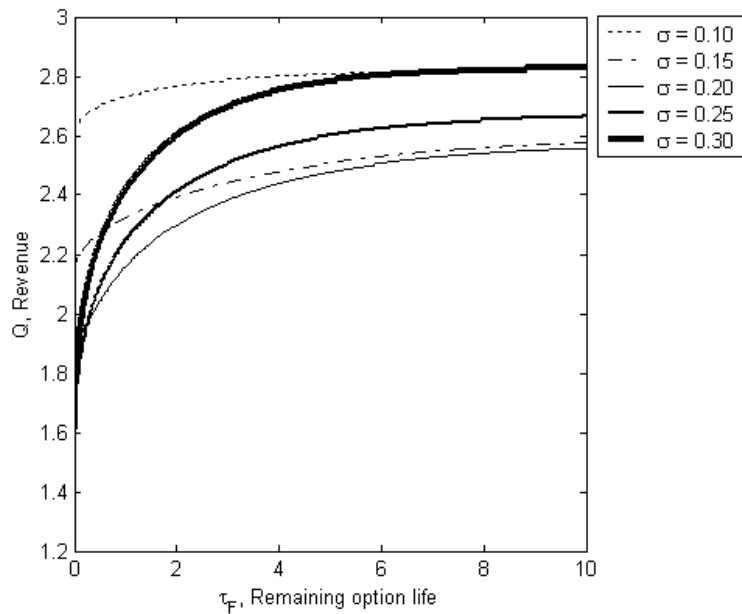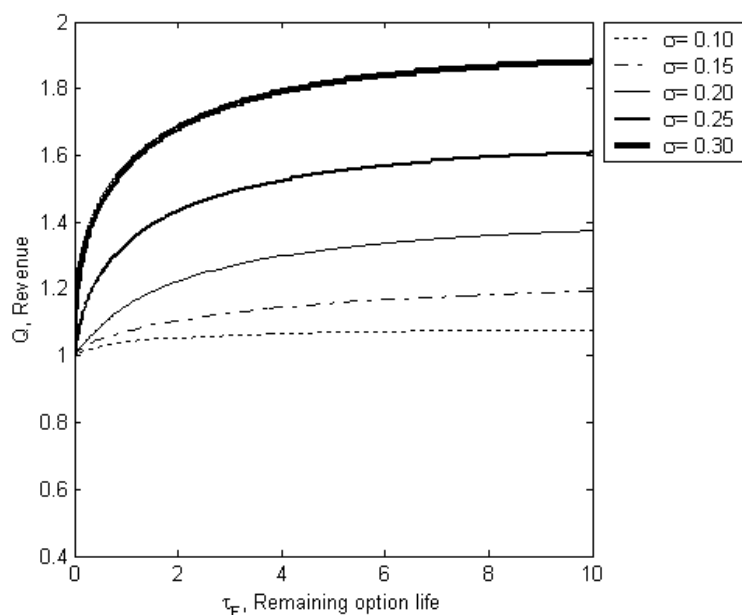FIGURE 4.3. A perpetual project and a finitely-lived option to invest: investment trigger boundary, $Q^*(\tau_F)$, for various levels of volatility and the set of parameters: $\mu = 0.08$, $r = 0.1$, $\rho = 0.7$, $\lambda = 0.4$, $I = 10$, $T = \infty$, $T_F = 10$.

to 0.15 moves the curve downwards.[7] But an increase of σ from 0.20 to 0.25 and 0.30 shifts the optimal triggers upwards. The important finding of this numerical analysis is that after comparing Figure 4.2 and Figure 4.1, we can conclude that the levels of σ at which the trigger decreases and increases with uncertainty, remain roughly the same. In both cases the revenue uncertainty level at which the change of sign occurs lies between $\sigma = 0.15$ and $\sigma = 0.20$. Thus the finite-life option assumption neither mitigates nor augments the positive relationship between investment and uncertainty due to the decreasing trigger.

Figure 4.2 shows also that the effect of uncertainty may differ depending on the remaining option life. The dashed curve of $\sigma = 0.15$ is below the dot-marked curve of $\sigma = 0.25$ at high τ_F and above at low τ_F. The reason is the nearly flat horizontal shape of the optimal investment trigger curve at relatively low σ ($\sigma = 0.10$ or $\sigma = 0.15$) for most of the option life and a sudden drop close to $\tau_F = 0$. This shape is caused by the convenience yield being low at lower σ, implying that there is only a small gain of undertaking the investment early (recall that a call option is never prematurely exercised if the convenience (dividend) yield is zero).

The behavior of the investment boundary in Figure 4.2 can be contrasted with the case of the perpetual project. Figure 4.3 shows that when the project life is infinite

---

[7]Except at the expiry date $\tau_F = 0$, at which $Q^*(t)$ increases in σ for all σ.

then $Q^*(t)$ moves upwards with increasing uncertainty. This is the usual monotonic relation consistent with the model with perpetual opportunity to invest.

4.4.2 Stochastic project life

An alternative for assuming a deterministic finite project life is to impose that a Poisson arrival brings the project to an end. We study this here and assume that the project lifetime (after installation) follows a Poisson process with rate γ. Among the numerous studies applying this set up we like to mention Merton (1976), who uses it in a financial option context, and McDonald and Siegel (1986), who apply it to the case of real investments.

Using equation (4.4) and the probability density of the stochastic lifetime, we obtain the project value

$$V(Q) = \int_0^\infty Q \frac{1 - e^{-(r + \lambda\rho\sigma - \mu)t}}{r + \lambda\rho\sigma - \mu} \gamma e^{-\gamma t} dt = \frac{Q}{r + \lambda\rho\sigma - \mu + \gamma}.$$

Note that the mortality rate γ leads to an environment equivalent to the one with perpetual projects except that the effective discount rate is now $r + \lambda\rho\sigma - \mu + \gamma$ rather than $r + \lambda\rho\sigma - \mu$. The resulting formula prompts that a project with stochastic lifetime can be interpreted as a perpetual project that is exponentially depreciated with rate γ (see Dixit and Pindyck (1994, p.200)).

Analogous to the previous analyses, the optimal investment trigger can be derived:

$$Q^* = \frac{\beta_1}{\beta_1 - 1} \left(r + \lambda\rho\sigma - \mu + \gamma \right) I. \tag{4.12}$$

We can now show that the non-monotonic uncertainty effect carries over to the case of a stochastic project life.

Proposition 4.6 *If $\gamma > 0$ and $\lambda\rho > 0$, then the uncertainty effect on the investment trigger is non-monotonic: it decreases in σ for low levels of σ and then increases. The length of the σ-interval where the negative effect occurs increases in γ.*

This result points out how strongly the monotonic relationship between the investment trigger and uncertainty hinges on the assumption of the project being perpetual. If there exists just a small probability that the project will be finished in finite time, the investment trigger will be decreasing with increasing uncertainty for a small enough σ. To illustrate this result, a numerical example is presented in Figure 4.4. Here we indeed see that even a very small γ causes the trigger to decrease in uncertainty at low but realistic levels of uncertainty. We also see that the boundary moves upward as
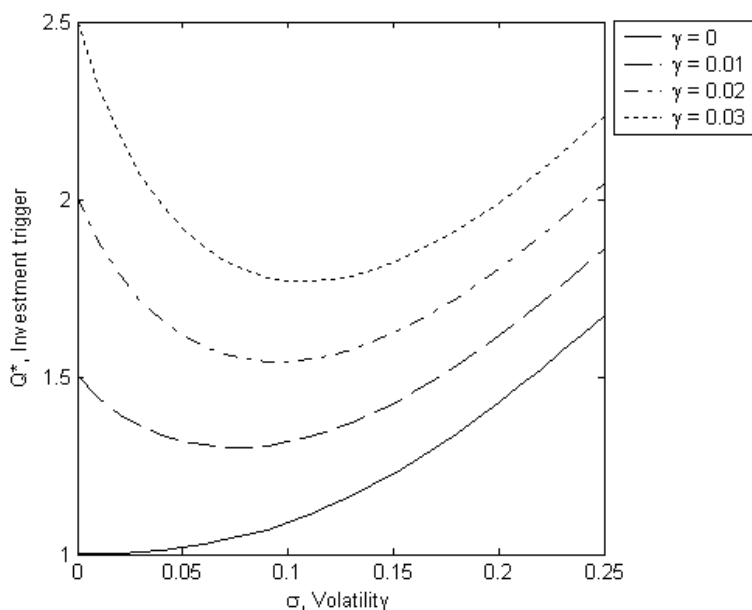
FIGURE 4.4. Investment trigger as a function of volatility for various Poisson arrival rates γ and the set of parameters: $\mu = 0.08$, $r = 0.1$, $\rho = 0.7$, $\lambda = 0.4$, $I = 10$.

γ increases, reflecting that a higher instantaneous flow is needed for the investment to be optimal, once the probability that a project ends increases.

4.4.3   General convenience yield

The previous results stated in Propositions 4.1-4.6 are obtained for the framework of Section 4.2 (and Section 4.4.2 in the stochastic life case). In that model, the equilibrium discount rate, and also the convenience yield, are determined by the standard CAPM and thus are linear in σ. Here we check whether this linearity is crucial for the results that we obtained. This issue is relevant as, apart from the standard CAPM, there exist theory and some evidence in favour of nonlinearity. For example, it is well-know that the presence of finite heterogeneous investment horizons leads to a non-linear CAPM with a nonlinear relationship between returns and risk (see, e.g., Lee, Wu and Wei (1990)). Moreover, there is a growing literature on factor pricing models with nonlinearities (see Bansal and Viswanathan (1993)).

Let the convenience yield be a non-decreasing, continuous, twice differentiable function of uncertainty, $\delta(\sigma)$ for $\sigma \geq 0$. In the previous sections we obtained results for the linear case, i.e. $\delta''(\sigma) = 0$. We now present propositions that generalize those results. Corresponding to Proposition 4.1 we have the following.

Proposition 4.7 *If the project life is infinite and $\delta'(\sigma) \geq 0$, then the investment trigger increases with uncertainty.*

Proposition 4.4 can be generalized as follows.

Proposition 4.8 *If the project life is finite, $\delta'(\sigma) > 0$, and $\delta''(\sigma) \leq 0$, then the uncertainty effect on the investment trigger is non-monotonic: it decreases in σ for low levels of σ and then increases. The length of the σ-interval where the negative effect occurs decreases with project lifetime.*

So in the case of a finite project life, the previously observed properties for linear $\delta(\sigma)$ carry over to a concave $\delta(\sigma)$. In case of a convex $\delta(\sigma)$, we can have either a U-shaped relationship or a monotonic one.[8]

4.4.4 Mean reverting revenues

In this section we relax the assumption that revenue follows a geometric Brownian motion by allowing Q to be mean reverting. There have been several studies that considered the impact of mean revision on real options valuation (Metcalf and Hassett (1995), Schwartz (1997), Sarkar (2003)). We analyze here whether our result that a finite project life may cause a non-monotonic investment-uncertainty relationship carries over to the framework with mean revision.

Suppose that the revenue flow follows a geometric mean reverting process characterized by the following stochastic differential equation:

$$dQ_t = [\mu Q_t + \kappa(\theta e^{\mu t} - Q_t)]dt + \sigma Q_t dZ_t. \tag{4.13}$$

The process corresponds to the generalized mean revision in equation (2) of Metcalf and Hassett (1995). $\kappa > 0$ is the speed of revision of the process towards its mean. The mean is $\theta e^{\mu t}$ and grows exponentially at rate $\mu > 0$. If $\kappa = 0$ the process becomes a geometric Brownian motion with drift μ as in (4.1). If $\mu = 0$, the process in (4.13) becomes a simple mean revision with constant mean as studied by Sarkar (2003).

Denote the project value with remaining time τ to maturity at time t by $V(Q, \tau, t)$ (the mean of Q depends on calendar time and this dependence is reflected in V). Using standard arguments, we find that $V(Q, \tau, t)$ must satisfy the following differential equation

$$\frac{1}{2}\sigma^2 Q^2 V_{QQ} + [(\mu - \lambda\rho\sigma)Q + \kappa(\theta e^{\mu t} - Q)]V_Q - V_\tau - rV + Q = 0, \tag{4.14}$$

---

[8]To check it, take, for instance, $\delta(\sigma) = r + \lambda\rho\sigma^{3/2} - \mu$ with the parameter values as in Table 4.1 and the uncertainty effect is U-shaped. However, if $\delta(\sigma) = r + \lambda\rho\sigma^3 - \mu$, the effect of uncertainty is always positive.

with the terminal condition at maturity $\tau = 0$

$$V(Q, 0, t) = 0. \tag{4.15}$$

Differential equation (4.14) with boundary condition (4.15) has an analytical solution[9] which is linear in Q:

$$V(Q, T, t) = AQ + B,$$

where

$$
\begin{aligned}
A &= \frac{1 - e^{-(r + \lambda\rho\sigma - \mu + \kappa)T}}{r + \lambda\rho\sigma - \mu + \kappa}, \\
B &= \frac{\kappa\theta e^{\mu t}}{\kappa + \lambda\rho\sigma}\left[\frac{1 - e^{-(r-\mu)T}}{r - \mu} - \frac{1 - e^{-(r + \lambda\rho\sigma - \mu + \kappa)T}}{r + \lambda\rho\sigma - \mu + \kappa}\right].
\end{aligned}
$$

As expected, when $\kappa = 0$ the value function is identical to (4.4) with revenues following a geometric Brownian motion. When $\mu = 0$ the formula simplifies to the value function in equation (2) in Sarkar (2003).

Similarly, using standard arguments one can show that the value of the option to invest $F(Q)$ satisfies

$$\frac{1}{2}\sigma^2 Q^2 F_{QQ} + [(\mu - \lambda\rho\sigma)Q + \kappa(\theta e^{\mu t} - Q)]F_Q - rF = 0, \tag{4.16}$$

with boundary conditions: $F(Q^*) = AQ^* + B - I$, $F_Q(Q^*) = A$ and $F(0) = 0$. The differential equation (4.16) with the boundary conditions has no known analytical solution, but it can be readily solved numerically. To find the optimal investment trigger we use a simple shooting method. The method is very accurate as long as the value function does not have to be evaluated numerically (see Dangl and Wirl (2003) for more details and further discussion). We convert the second order differential equation (4.16) into a system of two first order differential equations and employ a Runge–Kutta algorithm to solve the initial value problem.

To examine the effect of uncertainty on investment in the presence of mean revision, we repeat the numerical exercise for various project durations and levels of speed of revision κ. Figure 4.5 illustrates the results for two different project lifetimes $T = 10$ and 30, and various levels of κ. The other parameters are as in the previous numerical examples with the addition of $\theta = 0.5$ and $t = 0$. In principle, for each t the mean of Q is different (it grows deterministically and equals $\theta e^{\mu t}$) and so the trigger strategy

---

[9] The analytical solution for the project value with finite lifetime when revenues follow a generalized geometric mean reverting process (4.13) might be of interest on its own; see also Li (2003) who solves a similar problem.
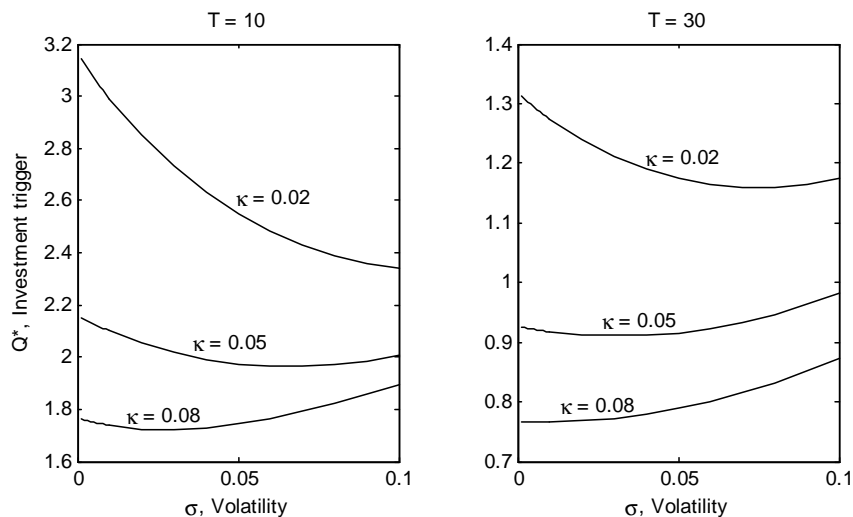
FIGURE 4.5. Investment trigger as a function of volatility for various levels of speed of revision κ and the set of parameters: $\mu = 0.08$, $r = 0.1$, $\rho = 0.7$, $\lambda = 0.4$, $\theta = 0.5$, $I = 10$, $t = 0$. Project life time T is 10 (left) and 30 (right).

changes over time. θ is the mean at $t = 0$ and its value is chosen in such a way that it is not above the optimal investment triggers at $t = 0$.

It is clear that in general uncertainty effects are less pronounced in the presence of mean revision. Therefore, as illustrated in Figure 4.5 uncertainty effects on investment are flattened especially for larger κ and long-lived projects. Yet the main result of this chapter still holds, since the non-monotonic relationship between uncertainty and investment is present if the project life is short and the region of the negative effect is larger the shorter is the project lifetime. For higher levels of κ and for larger T the uncertainty effect weakens and ultimately the effect only holds for very low values of σ.

4.5   Conclusions

Our analysis shows that a finite life of an investment project in combination with a risk premium in expected rates of return may reverse the usual effect of uncertainty on irreversible investments. In particular, we determined a scenario under which increased uncertainty reduces the value of waiting with investment. We now briefly discuss some implications of this result.

In corporate practice investment projects are usually considered to have a finite life, which supports the importance of our result. It thus seems that assuming the project life to be infinite, which is done in the overwhelming majority of real options contributions, is useful for simplicity reasons but dangerous since adverse uncertainty effects are lost.

From a policy point of view our results demonstrate that there exists a positive level of uncertainty at which the investment trigger admits its lowest value. If the policy aim is to increase investment, then the implication is that it is not necessarily optimal in all cases to decrease the level of uncertainty of policy instruments. However, any specific recommendation may be a bit far-reaching in the current single-firm model with a general source of uncertainty. To derive policy implications out of our non-monotonic investment-uncertainty relationship deserves a separate study. Similarly, in order to focus on the main features of the described mechanism, we have not attempted to construct a richer model of industry equilibrium. This can be done by considering a competitive industry (as in Leahy (1993) and Caballero and Pindyck (1996)) or imperfect competition (as in Smets (1991), Smit and Ankum (1993), Grenadier (1996) and Smit and Trigeorgis (2004)). However, we are quite confident that, qualitatively spoken, our result carries over to these frameworks.

Our non-monotonicity result accords with empirical findings of Bo and Lensink (2005). In a panel of Dutch firms, the investment-uncertainty relationship is positive at low levels of uncertainty and negative at high levels. Until now, a clear theoretical explanation for such empirical results is missing. The factors hastening investment with greater uncertainty indicated in this chapter lend themselves to empirical tests.

4.A   Appendix: Proofs

4.A.1   Deterministic project life

The derivative of the investment trigger (given in (4.6)) with respect to σ is

$$\frac{dQ^*}{d\sigma} = \frac{I\beta_1}{(\beta_1 - 1)^2} \frac{1}{\sigma^2\left(\beta_1 - \frac{1}{2}\right) + \mu - \lambda\rho\sigma} \frac{1}{1 - e^{-(r+\lambda\rho\sigma-\mu)T}} (M - N\Delta), \qquad (4.17)$$

where

$$M = (\beta_1 - 1)\left(\beta_1 + \frac{1}{2}\right)\lambda\rho\sigma^2 + (\beta_1 - 1)(r - \mu)\sigma + \beta_1(\mu - \lambda\rho\sigma)\lambda\rho - r\lambda\rho,$$

$$N = (\beta_1 - 1)\left(\beta_1 - \frac{1}{2}\right)\lambda\rho\sigma^2 + (\beta_1 - 1)(\mu - \lambda\rho\sigma)\lambda\rho,$$

$$\Delta = (r + \lambda\rho\sigma - \mu)T\left[e^{(r+\lambda\rho\sigma-\mu)T} - 1\right]^{-1}.$$

Denote the term $M - N\Delta$ by L_1. The first three fractions of (4.17) are always positive (recall that $\sigma^2\left(\beta_1 - \frac{1}{2}\right) + \mu - \lambda\rho\sigma = \partial L_0/\partial\beta\big|_{\beta=\beta_1} > 0$, as the derivative is evaluated at the higher root of the convex quadratic L_0). The sign of L_1 thus determines the

sign of the derivative. From (4.7) we observe that

$$(\mu - \lambda\rho\sigma)\,\beta_1 = -\frac{1}{2}\beta_1^2\sigma^2 + \frac{1}{2}\beta_1\sigma^2 + r,$$

which can be substituted twice into M and N to obtain

$$M = \frac{1}{2}\left(\beta_1 - 1\right)^2 \lambda\rho\sigma^2 + \left(\beta_1 - 1\right)\left(r + \lambda\rho\sigma - \mu\right)\sigma, \qquad (4.18)$$

and

$$N = \frac{1}{2}\left(\beta_1 - 1\right)^2 \lambda\rho\sigma^2 + \left(r + \lambda\rho\sigma - \mu\right)\lambda\rho. \qquad (4.19)$$

Proof of Proposition 4.1. First, suppose that $\lambda\rho > 0$. Combining $T \to \infty$ with (4.17) and (4.18), we obtain that

$$
\begin{aligned}
\frac{dQ^*}{d\sigma} &= \frac{I\beta_1}{\left(\beta_1 - 1\right)^2} \frac{1}{\sigma^2\left(\beta_1 - \frac{1}{2}\right) + \mu - \lambda\rho\sigma}\left[\left(\beta_1 - 1\right)\sigma\left(r + \frac{1}{2}\left(\beta_1 + 1\right)\lambda\rho\sigma - \mu\right)\right] \\
&> \frac{I\beta_1}{\left(\beta_1 - 1\right)^2} \frac{1}{\sigma^2\left(\beta_1 - \frac{1}{2}\right) + \mu - \lambda\rho\sigma}\left[\left(\beta_1 - 1\right)\sigma\left(r + \lambda\rho\sigma - \mu\right)\right] \\
&\geq 0,
\end{aligned}
$$

where the first inequality stems from the observation that $\frac{1}{2}(\beta_1 + 1) > 1$ and the second from the assumption that $r + \lambda\rho\sigma - \mu = \delta > 0$.

The two other possibilities $\lambda\rho = 0$ and $\lambda\rho < 0$ are covered by the proofs of Propositions 4.2 and 4.3, respectively. ■

Proof of Proposition 4.2. Within our model we can impose absence of the impact of systematic risk by setting $\rho = 0$. The derivative of the investment trigger (given in equation (4.6)) with respect to σ is

$$\frac{dQ^*}{d\sigma} = \frac{I\beta_1}{\left(\beta_1 - 1\right)^2} \frac{1}{\sigma^2\left(\beta_1 - \frac{1}{2}\right) + \mu} \frac{1}{1 - e^{-(r-\mu)T}}\left(\beta_1 - 1\right)\sigma\left(r - \mu\right).$$

The resulting expression is always positive if $r > \mu$, which holds by the assumption that $\delta > 0$. ■

Proof of Proposition 4.3. Suppose that $\lambda\rho < 0$. Then the assumption that $\delta > 0$ holds if and only if $\sigma \in [0, \bar{\sigma})$, where $\bar{\sigma} = \frac{\mu - r}{\lambda\rho}$. We have that, denoting $\delta(\cdot)$ and $\beta_1(\cdot)$ as functions of σ, $\delta(\bar{\sigma}) = 0$ and $\beta_1(\bar{\sigma}) = 1$. So $[0, \bar{\sigma})$ is the relevant domain for σ in this case. Next, we claim that

$$\frac{1}{2}\left(\beta_1 + 1\right)\sigma < \bar{\sigma}, \text{ for all } \sigma \in [0, \bar{\sigma}). \qquad (4.20)$$

To verify, note that

$$\frac{d}{d\sigma} \frac{1}{2} (\beta_1 + 1) \sigma = \left[\sigma^2 \left(\beta_1 - \frac{1}{2} \right) + \mu - \lambda\rho\sigma \right]^{-1} \left[\frac{1}{2} (3\beta_1 - 1) \sigma^2 + (\beta_1 - 1) \mu - \lambda\rho\sigma \right]$$
$$> 0$$

and $\frac{1}{2} (\beta_1(\bar\sigma) + 1) \bar\sigma = \bar\sigma$. So, for positive σ less than $\bar\sigma$, the inequality (4.20) is true.

Now, $\lambda\rho < 0$ implies that $N < 0$. Combining (4.20) and (4.18) we have that

$$M = (\beta_1 - 1) \sigma \left[r + \frac{1}{2} (\beta_1 + 1) \lambda\rho\sigma - \mu \right]$$
$$> (\beta_1 - 1) \sigma (r + \lambda\rho\bar\sigma - \mu) = (\beta_1 - 1) \sigma\delta(\bar\sigma) = 0.$$

Since $M > 0$, $N < 0$, and $1 \geq \Delta > 0$, the derivative (4.17) is also positive and the proposition is proved. ∎

Proof of Proposition 4.4. Suppose that T is finite and $\lambda\rho > 0$. We want to show that L_1 is negative for low $\sigma \geq 0$ and becomes positive when σ increases. First, it is useful to observe the simple fact that $1 \geq \Delta > 0$ and

$$\frac{d\Delta}{d\sigma} < 0. \tag{4.21}$$

It can also be verified that

$$L_1 \leq 0 \Rightarrow (\beta_1 - 1) \sigma - \lambda\rho < 0 \iff \frac{d\beta_1}{d\sigma} > 0. \tag{4.22}$$

Then note that at $\sigma = 0$, $L_1 = - (r - \mu) \lambda\rho\Delta < 0$. So $\frac{dQ^*}{d\sigma}$ is also negative at $\sigma = 0$. As σ increases, Δ converges to zero and L_1 becomes positive. We show now that L_1 changes its sign from negative to positive only once with increasing σ. If $L_1 = 0$, then $\Delta = \frac{M}{N}$ and

$$\frac{dL_1}{d\sigma} = \frac{dM}{d\sigma} - \frac{dN}{d\sigma}\Delta - N\frac{d\Delta}{d\sigma} > \frac{dM}{d\sigma} - \frac{dN}{d\sigma}\Delta = \frac{1}{N} \left(\frac{dM}{d\sigma}N - \frac{dN}{d\sigma}M \right)$$
$$= \frac{\delta\lambda\rho}{N} \left[\frac{d\beta_1}{d\sigma}\sigma + \beta_1 - 1 \right] \{ (\beta_1 - 1) [\lambda\rho - (\beta_1 - 1) \sigma] \sigma + \delta \}$$
$$> 0. \tag{4.23}$$

The inequalities follow from (4.21) and (4.22). So L_1 increases in σ at the point at which $L_1 = 0$. Now, continuity of L_1 implies that it changes its sign only once from negative to positive at some $\sigma^* > 0$. Hence the first part of the proposition is proved.

To verify that the σ-interval where the negative effect occurs is larger the shorter is the project life, we consider

$$\frac{d\sigma^*}{dT} = -\left.\frac{\frac{\partial L_1}{\partial T}}{\frac{\partial L_1}{\partial \sigma}}\right|_{\sigma=\sigma^*} = \left.\frac{N\frac{d\Delta}{dT}}{\frac{\partial L_1}{\partial \sigma}}\right|_{\sigma=\sigma^*} < 0.$$

The inequality follows from the fact that $\frac{d\Delta}{dT} < 0$ and (4.23). ∎

Proof of Proposition 4.5. The sum of the two option effects is

$$\frac{\partial Q^*}{\partial \beta_1}\frac{\partial \beta_1}{\partial \sigma} + \frac{\partial Q^*}{\partial \beta_1}\frac{\partial \beta_1}{\partial \delta}\frac{\partial \delta}{\partial \sigma} = \frac{I\beta_1}{(\beta_1 - 1)^2}\frac{\delta(\sigma)}{1 - e^{-\delta(\sigma)T}}\frac{(\beta_1 - 1)\sigma - \lambda\rho}{\sigma^2\left(\beta_1 - \frac{1}{2}\right) + \mu - \lambda\rho\sigma}. \qquad (4.24)$$

As $\beta_1 > 1$ and $\sigma^2\left(\beta_1 - \frac{1}{2}\right) + r - \delta(\sigma) > 0$, the sign of expression (4.24) depends on the sign of $L_2 \equiv (\beta_1 - 1)\sigma - \lambda\rho$ in the way stated in the proposition.

It remains to be shown that there exists a unique non-negative $\hat{\sigma}$. Note that, if $\lambda\rho > 0$, at $\sigma = 0$ we have that $L_2 = -\lambda\rho < 0$ and the combined option effect is negative. To verify that the option effect changes its sign only once from negative to positive with increasing σ, we show that L_2 (being continuous in $\sigma > 0$) always increases with σ if $L_2 \leq 0$. That is,

$$\frac{dL_2}{d\sigma} = \frac{\lambda\rho\sigma - (\beta_1 - 1)\sigma^2}{\sigma^2(\beta_1 - 1) + \mu - \lambda\rho\sigma} + \beta_1 - 1 \geq \beta_1 - 1 > 0,$$

if $L_2 \leq 0$.

The discounting effect is given by

$$\frac{\partial Q^*}{\partial \delta}\frac{\partial \delta}{\partial \sigma} = \frac{I\beta_1}{\beta_1 - 1}\frac{1 - e^{-\delta(\sigma)T} - \delta(\sigma)Te^{-\delta(\sigma)T}}{\left(1 - e^{-\delta(\sigma)T}\right)^2}\lambda\rho,$$

which is always positive and increasing in T. It is straightforward from derivations leading to (4.24) that $\frac{\partial Q^*}{\partial \beta_1}\frac{\partial \beta_1}{\partial \sigma}$ and $\frac{\partial Q^*}{\partial \beta_1}\frac{\partial \beta_1}{\partial \delta}\frac{\partial \delta}{\partial \sigma}$ decrease in absolute terms in T. ∎

4.A.2 Stochastic project life

Let $\delta(\sigma)$ be a continuous twice differentiable convenience yield function. The derivative of Q^* given in (4.12) with respect to σ eventually becomes:

$$\frac{dQ^*}{d\sigma} = \frac{I\beta_1}{(\beta_1 - 1)^2}\frac{1}{\sigma^2\left(\beta_1 - \frac{1}{2}\right) + r - \delta(\sigma)}L_3, \qquad (4.25)$$

where

$$L_3 = \frac{1}{2} (\beta_1 - 1)^2 \delta'(\sigma)\sigma^2 + (\beta_1 - 1) \delta(\sigma)\sigma + [(\beta_1 - 1)\sigma - \delta'(\sigma)]\gamma. \qquad (4.26)$$

The first two fractions of the right-hand side of (4.25) are always positive, so the sign of the derivative is determined by the sign of L_3.

Proof of Proposition 4.6. The proof follows from the proof of Proposition 4.8 below with linear $\delta(\sigma)$. ∎

We prove Propositions 4.7 and 4.8 only for stochastic project lifetime; similar proofs can be obtained for the deterministic case.

Proof of Proposition 4.7. Note that if $\gamma = 0$ and $\delta'(\sigma) > 0$ then
$L_3 = \frac{1}{2} (\beta_1 - 1)^2 \delta'(\sigma)\sigma^2 + (\beta_1 - 1) \delta(\sigma)\sigma > 0.$ ∎

Proof of Proposition 4.8. We want to show that for $\gamma > 0$, $\delta'(\sigma) > 0$ and $\delta''(\sigma) < 0$, L_3 is negative for low $\sigma \geq 0$ and turns to positive with increasing σ. First we note that at $\sigma = 0$, $L_3 = -\delta'(0)\gamma < 0$. Then observe that a straightforward consequence of (4.26) is that

$$L_3 \leq 0 \Rightarrow (\beta_1 - 1)\sigma - \delta'(\sigma) < 0 \iff \frac{d\beta_1}{d\sigma} > 0. \qquad (4.27)$$

Using this, if $L_3 \leq 0$, we have that

$$\begin{aligned}
\frac{dL_3}{d\sigma} &= \frac{d\beta_1}{d\sigma} \left[(\beta_1 - 1) \delta'(\sigma)\sigma^2 + \sigma (\delta(\sigma) + \gamma) \right] + (\beta_1 - 1)^2 \delta'(\sigma)\sigma \\
&\quad + (\beta_1 - 1) (\delta(\sigma) + \gamma + \delta'(\sigma)\sigma) + \left[\frac{1}{2} (\beta_1 - 1)^2 \sigma^2 - \gamma \right] \delta''(\sigma) \\
&> \left[-\frac{1}{\delta'(\sigma)} (\beta_1 - 1) \sigma (\delta(\sigma) + \gamma) \right] \delta''(\sigma) > 0.
\end{aligned}$$

So L_3 always increases in σ if $L_3 \leq 0$. From the continuity of L_3 it now follows that L_3 changes its sign only once from negative to positive at some $\sigma^* > 0$. This proves the first part of proposition.

To verify that the σ-interval where the negative effect occurs is larger, the shorter is the project life we consider

$$\frac{d\sigma^*}{d\gamma} = -\frac{\frac{\partial L_3}{\partial \gamma}}{\frac{\partial L_3}{\partial \sigma}}\bigg|_{\sigma=\sigma^*} = \frac{\delta'(\sigma) - (\beta_1 - 1)\sigma}{\frac{\partial L_3}{\partial \sigma}}\bigg|_{\sigma=\sigma^*} > 0,$$

where for the inequality we employ (4.27) and the first part of the proof of this proposition. ∎

5

Partial Divestment and Firm Sale under Uncertainty

5.1 Introduction

Firms can downgrade their operations and release the capital to investors in response to unfavorable market conditions or a deterioration of efficiency relative to competitors. In essence, corporate assets can be either divested and sold gradually over time or the whole firm can be sold at once. These two alternative phase-out modes differ in two key aspects. On the one hand, gradual divestment allows firms to maintain flexibility and to benefit from possible future positive market developments. In this respect gradual divestment is advantageous compared to firm sale. On the other hand, partial displaced assets are sold with a discount on secondary markets whereas firms are sold with a substantial takeover premium. In this chapter we study optimal divestment directly addressing the trade-off between the flexibility of gradual divestment and the premium of whole firm sale.

The flexibility advantage of gradual divestment is related to the optionality of the irreversible (dis-)investment decisions. The real options approach to investment stresses the value created by uncertainty when investment timing is flexible. In the case of gradual divestment, the firm holds a bundle of options to sell its partial assets. A marginal sale of assets leaves the options to sell the remaining assets and allows the firm to benefit from their optimal execution in the future. In the case of firm sale, the decision is also an option at owners discretion. The available evidence on takeover transactions supports the stance we adopt in this chapter. Andrade, Mitchell and Stafford (2001) show that 94 percent of takeover transactions are initiated by the

selling party.[1] While the timing of firm sales is flexible, all flexibility is lost after the firm sale and exit.

If the whole firm is sold at the same price as the sum of partial asset sales, gradual divestment is always a preferable choice. This is no longer the case if partial asset sale is associated with a discount in comparison to whole firm sale. The literature on asset sale provides strong empirical evidence for the partial asset sale discount and the firm sale premium. The discount for partial displaced capital stems from firm and sectorial capital specificity, the thinness of the used capital market and costs of redeploying the capital. For example, Ramey and Shapiro (2001) cite such reasons for substantially discounted prices of used capital relative to replacement value found in the aerospace industry. Pulvino (1998) shows that financial constraints add to depress selling prices for used aircraft in transactions between airlines. Firm sales, on the other hand, are attributed with premiums relative to some benchmark values. The two main sources of the premium are the following. First, a firm is sold with a premium over the selling price of partial physical capital because many types of intangible assets are purchased only with the full corporate entity. These assets include mainly competitive intangibles such as customer and suppliers relations, know-how and organization, and may account to a significant portion of firm value (see, e.g., Hand and Lev (2003)). Second, persistent empirical evidence documents substantial takeover premiums defined as the difference between the selling price and the value of the target firm before the transaction. A recent study of Boone and Mulherin (2007) reports a mean premium of 40 percent in the announced transaction price relative to the price of the target firm 4 weeks before the first announcement of the takeover. This means that even after controlling for intangible assets (included in the pre-announcement firm value), whole firms are sold with premiums. These takeover premiums are typically explained as originating from strategic synergies or higher productivity of the buying firm coupled with bargaining power of the selling party. Part of the surplus created by a merger is paid out to the target firm owners.

Given the above characteristics of corporate divestments, some interesting questions remain unanswered. What does the optimal downsizing path look like? How does the structure of the price discount-premium affect the choice between partial divestment and firm sale? Should firms with more volatile profits divest partially or sell at once?

---

[1] Using a smaller sample, but with more detailed information, Boone and Mulherin (2007) document that 15 percent of takeover bids are unsolicited. However small is the fraction of unsolicited takeover bids, even these sale transactions leave some flexibility and discretion in the hands of the selling party. Boone and Mulherin (2007) report that most of the unsolicited bids are executed by competitive auctions to solicit bids from other potential buyers. Furthermore, Schwert (2000) shows that the so-called hostile takeover deals are economically equivalent to friendly takeovers and hostility is mostly related to strategic negotiations.

Do firms in more declining markets prefer gradual divestment or firm sale? Do firms with more industry-specific capital opt for gradual exit or takeover sale?

To answer these questions we construct a stylized real options model in which a firm faces a stochastic profit flow and optimally chooses its divestment path. Marginal units of capital may be released and sold at a discounted unit price. Alternatively, the whole firm can be sold at a premium price that depends on the capital level at the transaction time. To focus on the main trade-off problem between partial divestment and firm sale we assume that both decisions are irreversible. From a technical point of view, the problem is not trivial as it involves two different stochastic control technics. Partial divestment is modeled as a barrier control, and the firm adjusts capital level at each time the underlying profitability state variable reaches a new minimum on a barrier. On the other hand, whole-firm sale is a discrete decision, and the firm's problem takes the form of an optimal stopping problem.[2]

Our analysis indicates that the optimal divestment policy depends critically on the structure of the discount-premium of the capital price. We first study the simplest case, in which the firm-sale premium is linear (proportional in the level of capital). In this case, the optimal policy is either to divest only gradually if the proportional premium is below a certain threshold or to divest the whole firm if the proportional premium is sufficiently large (it is assumed here that the firm has followed the optimal divestment path before and does not start off the optimal policy path).

The optimal divestment policy takes a notably different form if the firm-sale premium is affine, i.e. if the premium consists of both proportional and fixed components. The fixed part of the premium arises because of, e.g., non-tangible assets sold only with the whole firm. In this case, if the proportional premium is sufficiently large, the firm optimally decides to use only the firm-sale option, as the premium offsets the gains from the flexibility of gradual divestment. But if the proportional premium is not too high, the firm optimally divests marginal units of capital in a declining market until its size reaches a certain threshold. Subsequently, the remaining capital is sold with the whole firm, but this only happens after an anticipation phase in which the market falls to a sufficiently low level. Intuitively, while at high levels of capital the firm prefers to maintain the flexibility of partial divestment against a moderate firm-sale premium, at lower levels of capital the benefit of a positive fixed premium will offset the flexibility advantage of gradual adjustments.

The model generates some new predictions on the optimal choice of divestment policy and, specifically, on the choice between partial divestment and firm sale. We

---

[2] Two other recent papers study corporate investment as mixed stochastic control problems. Guo and Pham (2005) analyzes optimal entry and subsequent investment, and Décamps and Villeneuve (2007) deals with dividend choice and optimal exercise of a growth option of a financially constrained firm.

find that in more uncertain markets the value-maximizing firm is more inclined to divest its capital fully at once. This means that, somewhat surprisingly, the value of flexibility of partial divestment does not become more valuable in more volatile markets compared to one-time firm sale. This effect arises because firm sale, being less flexible, has a higher value of waiting, which is directly and positively affected by uncertainty. We also show that firm sale is more preferable over partial divestment in more declining markets. This is because in a declining market there is less room to benefit from the flexibility of gradual divestment.

We extend the model to allow the selling price of capital to be correlated with the market state variable. The correlation coefficient between the market state and the price level is interpreted as a measure of industry-specificity of capital. We model in a reduced form the effect that, in a declining market, the demand for used capital decreases, and consequently prices also fall. We are interested how the industry-specificity of capital affects optimal divestment policies. We obtain that the more industry-specific is capital, the more preferable is partial divestment over firm sale. The explanation for this result is again related to the large value of waiting in the option to sell the firm at once. Because the specificity of capital affect the values of alternative strategies mostly via the values of waiting, and increasing specificity decreases these values, firm sale becomes less desirable.

The distinction between incremental capital adjustment and full-firm sale has been noted by several previous authors. In a series of two papers Ghemawat and Nalebuff (1985, 1990) study divestment and exit in declining industries. Ghemawat and Nalebuff (1985) consider the equilibrium order of full-firm exit in an oligopolistic market, while Ghemawat and Nalebuff (1990) allows firms to adjust their capital incrementally. In contrast, our paper incorporates both modes of capital adjustment in one model with uncertain demand, but we choose not to focus on the competitive effects. Lieberman (1990) and Maksimovic and Phillips (2001) empirically study the choice between partial and whole-firm divestment. While these studies do not test the whole set of predictions implied by our model, they nevertheless provide some supporting evidence. In particular, Lieberman (1990) and Maksimovic and Phillips (2001) show that large firms adjust capital partially and small firms tend to sell their all assets at once.

The remainder of the chapter is organized as follows. In Section 5.2 we set up a model of a firm with both partial and full-firm divestments. Section 5.3 derives the optimal divestment policies and the corresponding firm values. Section 5.4 discusses the implications of the model for divestment policies. Section 5.5 studies the effects of industry-specificity of capital. Section 5.6 concludes and the Appendix provides the proofs omitted in the main text.

5.2   Model

Consider a firm that produces a uniform non-storable good and faces stochastic demand. To produce the good the firm uses capital and possibly other variable inputs. The firm's operating profit at time t depends on the installed capital stock K_t and the market conditions variable X_t and is given by

$$\pi_t = \pi(X_t, K_t) = X_t K_t^{\gamma}, \quad \gamma \in (0,1). \tag{5.1}$$

The formulation has been frequently employed in previous studies (Bertola and Caballero (1994), Abel and Eberly (1996), Abel and Eberly (1999), Guo et al. (2005)) and is consistent with either a monopolist facing an isoelastic demand function and production technology with non-increasing returns to scale, or a price taking firm with decreasing returns to scale technology.[3] The investors are risk neutral and discount cash flows at a constant rate r.

The market conditions variable X_t captures the exogenous time varying business environment; more specifically X_t reflects demand shocks, but can also include productivity shocks and the prices of variable inputs (see footnote 3). We assume that the process $(X_t)_{t \geq 0}$ evolves according to the geometric Brownian motion

$$dX_t = \mu X_t dt + \sigma X_t dZ_t,$$

where Z_t is the standard Brownian motion, μ is the drift parameter and $\sigma > 0$ is the volatility parameter. We denote the filtration (the σ-algebra) generated by $(X_t)_{t \geq 0}$ with $(\mathcal{F}_t)_{t \geq 0}$. To ensure convergence of the problem, it is assumed that $\mu < r$.

Marginal units of capital can be sold at a price normalized to 1. Selling the whole firm at once results in a premium with a fixed component $A \geq 0$ and a unit price of capital equal to $a \geq 1$.[4] This means that the owners of the firm with a level of capital k divesting at once receive $ak + A$. The fixed premium may stem from the non-tangible assets or from a part of the takeover premium. It must be understood that our formulation incorporates the discount for partial displaced capital in the difference

---

[3]Suppose that the production function is $Q_t = K_t^{\phi}$, where Q_t is output produced at time t and $\phi \in (0,1]$ measures the degree of returns to scale. The inverse demand function is given by $P_t = X_t Q_t^{-\frac{1}{\varepsilon}}$, so that for a given quantity the price evolves in time together with the demand shock X_t. $\varepsilon > 1$ is the constant price elasticity of demand. Then instantaneous operating profit at time t is

$$\pi_t = P_t Q_t = X_t Q_t^{\frac{\varepsilon-1}{\varepsilon}} = X_t K_t^{\phi \frac{\varepsilon-1}{\varepsilon}}.$$

Defining $\gamma = \phi - \phi/\varepsilon$ we obtain (5.1) with $\gamma \in (0,1)$ if either the firm has a monopoly power (that is if $\varepsilon < \infty$) or the technology exhibits decreasing returns to scale ($\phi \in (0,1)$). As shown by Abel and Eberly (2004) the argument can be extended to the case with variable outputs in the production function (e.g. labor) and time varying productivity.

[4]The unit prices of capital are time constant in the current setup, but we relax this assumption in Section 5.5, where we allow for stochastic capital sale prices that are correlated with the market conditions variable.

between a and 1, so the normalization of the selling price of partial capital is without loss of generality. Capital divestment, either marginal or complete exit, is irreversible.

The objective of the firm is to maximize the value of the original owners. The control policy comprises the choice of capital and the exit time. The admissible capital contraction is a non-increasing process $K = (K_t)_{t \geq 0}$ that is progressively measurable with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$. The exit time τ is a stopping time with respect to $(\mathcal{F}_t)_{t \geq 0}$. The value of the firm following the optimal investment policy is the solution to the following optimization problem:

$$W(X_t, K_t) = \sup_{\tau} \sup_{\{K_{t+s}\}} \mathbb{E}_t \left[\int_0^{\tau - t} e^{-rs} \pi(X_{t+s}, K_{t+s}) ds \right.$$
$$\left. + \int_0^{\tau - t} e^{-rs} dK_{t+s} + e^{-r(\tau - t)} (aK_\tau + A) \right]. \quad (5.2)$$

The firm's problem involves two stochastic control problems, i.e. instantaneous control over a divestment path $\{K_{t+s}\}$ and optimal stopping at a stopping time τ.

5.3   Optimal divestment policy

5.3.1   Benchmark cases and linear premium

In this subsection we consider the two limit cases. In the first case, the firm has only gradual divestment available. In the second case, the firm can only downsize by firm sale. Both cases are straightforward simplifications of the more general optimization problem (5.2). This analysis is then used to study the case where both divestment modes are available and the firm-sale premium is linear in capital, i.e. $a \geq 1$ and $A = 0$.

Denote by $V^m(x, k)$ the value of the firm that follows optimal divestment policy in the case the firm can only sell partial capital. The optimal policy is characterized by a barrier function $X^m(k)$ that, for a given k, triggers infinitesimal divestment (see Pindyck (1988), Abel and Eberly (1996)). The standard arguments lead to the following Bellman equation that must be satisfied by V^m:

$$rV^m(x, k) = \frac{1}{2}\sigma^2 x^2 V_{XX}^m(x, k) + \mu x V_X^m(x, k) + \pi(x, k). \quad (5.3)$$

The equation states that the required rate of return (the left-hand side) must be equal to the expected gain in firm value plus profit flow $\pi(x, k)$ (the right-hand side).

The divestment trigger $X^m(k)$ and the value V^m can be obtained by solving the differential equation (5.3) subject to appropriate boundary conditions. At the divest-

ment trigger the firm sells the infinitesimal capital dk for a price of 1 per unit. It must hold that $V^m(X^m(k), k) = V^m(X^m(k), k - dk) + dk$. Writing this in derivative form, we obtain the smooth-pasting condition

$$V_K^m(X^m(k), k) = 1. \tag{5.4}$$

The condition requires that the marginal value of capital at the optimal divestment barrier $X^m(k)$ must be equal to its selling price.

The optimality condition for $X^m(k)$ requires that the slopes of the value function are equal at $X^m(k)$. The requirement in derivative form is known as the high-contact condition (see Dumas (1991)) and is written as

$$V_{XK}^m(X^m(k), k) = 0. \tag{5.5}$$

Finally, we also require that, as the demand shift increases to infinity, the option value to divest remains finite. This means that[5]

$$\lim_{x \to \infty} V^m(x, k) - \frac{\pi(x, k)}{r - \mu} < \infty. \tag{5.6}$$

In the second extreme case, the firm has only the option to phase out by firm sale. Denote by $V^e(x, k)$ the value function of the firm following an optimal firm sale policy at trigger $X^e(k)$. Given that the values in both cases are driven by the same stochastic process and the same payoff function, it is clear that before exit, V^e must satisfy the same type of Bellman equation as before:

$$rV^e(x, k) = \frac{1}{2}\sigma^2 x^2 V_{XX}^e(x, k) + \mu x V_X^e(x, k) + \pi(x, k). \tag{5.7}$$

In order to obtain the firm value and the optimal trigger strategy, we need to solve (5.7) subject to the appropriate boundary conditions. When the trigger $X^e(k)$ is reached, the firm sells k units of capital for unit price a and obtains a non-negative fixed premium A. The value function must be equal to the proceeds from sale, which means that the value-matching condition is

$$V^e(X^e(k), k) = ak + A. \tag{5.8}$$

The firm maximizes its value by choosing the optimal $X^e(k)$ and this requires that the slopes of the value function are equal at the sale trigger. This translates into the

---

[5] The discounted expected profit flow (the second term on the left-hand side) goes to infinity as $x \to \infty$, but the remaining value, i.e. the value of the option to divest, should be finite.

smooth-pasting condition at $X^e(k)$:

$$V_X^e\left(X^e(k), k\right) = 0. \tag{5.9}$$

In addition, the value function should be finite as X raises to infinity, so that the firm-sale option remains finite:

$$\lim_{x \to \infty} V^e(x, k) - \frac{\pi(x, k)}{r - \mu} < \infty. \tag{5.10}$$

Using the above analysis, we prove the first result of the mixed case where both gradual divestment and firm sale are available, and the firm sells at a proportional premium. Before we state the result, let us define a^* by

$$a^* = \frac{1}{\gamma}\left[\frac{1 - \beta\left(1 - \gamma\right)}{\gamma}\right]^{\frac{1}{\beta - 1}}.$$

Proposition 5.1 *Suppose that $a \geq 1$, $A = 0$ and (X_0, K_0) is at or above the relevant triggers characterized below.*
(a) If $a < a^*$, the firm divests only via partial divestment at

$$X^m(k) = \frac{\beta}{\beta - 1}\frac{1}{\gamma}\left(r - \mu\right)k^{1 - \gamma},$$

and the firm value is

$$W(x, k) = B_1(k)x^\beta + \frac{1}{r - \mu}xk^\gamma,$$

where

$$B_1(k) = \frac{1}{1 - \beta}\frac{k}{1 - \beta\left(1 - \gamma\right)}X^m(k)^{-\beta}$$

and

$$\beta = \frac{1}{2} - \frac{\mu}{\sigma^2} - \sqrt{\left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}} \leq 0.$$

(b) If $a \geq a^*$, then the firm sale trigger is given by

$$X^e(K_0) = \frac{\beta}{\beta - 1}a\left(r - \mu\right)K_0^{1 - \gamma}$$

and the firm value is

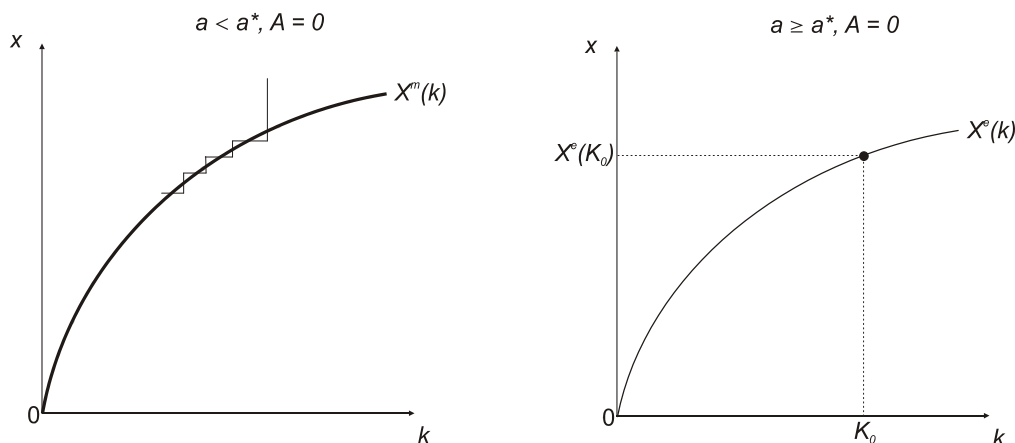$$W(x, k) = B_2(k)x^\beta + \frac{1}{r - \mu}xk^\gamma,$$

FIGURE 5.1. Divestment triggers with linear firm-sale premium. The left panel presents the case of $a < a^*$ and $A = 0$. In this case the firm divests only partially following barrier control at $X^m(k)$. The right panel presents the case $a \geq a^*$ and $A = 0$. In this case the firm divests only by firm sale at trigger $X^e(K_0)$.

where

$$B_2(k) = \frac{ak}{1 - \beta} X^e(k)^{-\beta}.$$

The proposition characterizes the optimal divestment triggers and the firm values in two cases. When the proportional premium is sufficiently large, $a \geq a^*$, the whole firm is sold at once as soon as the market shock reaches $X^e(K_0)$. If $a < a^*$, the firm divests only gradually following the barrier control at $X^m(k)$. Figure 5.1 presents the optimal divestment policies in the two cases. The reason for this dichotomous outcome is that the proportionality of payoffs in the two alternative divestment modes translates into the proportionality of the value function. If the premium is sufficiently small, then flexibility of partial divestment always offsets the premium of firm sale. If a is sufficiently large, then the premium counterbalances the flexibility advantage of partial divestment at all levels of capital.[6]

---

[6] The results and the conclusions presented here depend on the assumption that (X_0, K_0) is at or above the relevant triggers. The case is economically the most interesting. For the starting value to fall below the triggers, the firm must have deviated for some unmodeled reasons from the optimal policy before the initial date. Nevertheless, if $a < a^*$ and $X_0 \leq X^m(K_0)$ (in other words, the firm starts "too large" for a given market), the analysis resembles the model of Décamps, Mariotti and Villeneuve (2006) that studies an investment decision in one of two alternative projects. For a given x, there is a level of capital at which the firm is indifferent between partial divestment and whole-firm sale. Intuitively, if the firm has a high level of capital for the current (low) state of the market, it is better off selling all the capital with a premium than making a large partial adjustment at discounted prices and stay at the low market. If x falls below this indifference point, firm sale is preferable, if x rises, the value of partial divestment will exceed the value of firm sale. As in Décamps et al. (2006), it is possible to show that at the point of indifference the firm optimally does not make an divestment decision, and instead prefers to wait for the development of the market to decide for either partial adjustment, if x increases sufficiently, or firm sale, if x falls sufficiently and the market becomes unattractive for partial adjustment. The bottom line is that there is an inaction region at low levels of x for a given k, in which the firm does not make divestment decisions, but divest the whole firm if the market deteriorates and divests partially if the market *improves*.

5.3.2   Divestment with non-linear firm-sale premium

In this section we consider a more general case of firm-sale premium and allow it to be affine in the level of capital. In other words, we assume that $a \geq 1$ and $A > 0$. The previous section shows that with $A = 0$, $a \geq a^*$ implies that $V^e(x, k) \geq V^m(x, k)$ and the firm is better off selling the whole entity. As we show next, this conveys to the affine case, but if $a < a^*$, it needs no longer be true that $V^e(x, k) < V^m(x, k)$ for all levels of capital.

Lemma 5.2 *Suppose that $a \geq 1$ and $A > 0$. If $a \geq a^*$, then $V^e(x, k) \geq V^m(x, k)$. If $a < a^*$, then there exists a level of capital $\tilde{k}$ that separates two regimes: $V^e(x, k) \leq V^m(x, k)$ for $k \geq \tilde{k}$, and $V^e(x, k) > V^m(x, k)$ for $k < \tilde{k}$.*

In the affine case, $V^e(x, k)$ exceeds $V^m(x, k)$ for sufficiently low k. The intuition is that at small levels of capital the benefit of achieving a positive fixed premium will offset the flexibility advantage of partial divestment. However, the inequality $V^e(x, k) > V^m(x, k)$ is only a necessary condition for whole-firm sale. Even if $V^e(x, k) > V^m(x, k)$ holds, the firm may still be better off selling some capital by partial divestment before selling the remaining capital at once. This will be the case as long as the marginal value of partial divestment exceeds the marginal value of capital sold with the whole firm. These arguments suggest that in the case of $a < a^*$, optimal divestment will take the form of a two-stage policy. If the capital level is relatively large, such that it exceeds a certain threshold on capital K^*, the firm will optimally divest partially. Below K^*, investors will be better off selling the whole firm. The aim of the remainder of this section is to characterize this policy and the corresponding firm value.

As before, it is standard to show that the value function $W(x, k)$ satisfies the following Bellman equation:

$$rW(x, k) = \frac{1}{2}\sigma^2 x^2 W_{XX}(x, k) + \mu x W_X(x, k) + \pi(x, k). \qquad (5.11)$$

The optimal solution to the optimization problem (5.2) can be characterized using the differential equation (5.11) and the appropriate boundary conditions. As long as $k > K^*$, the marginal value of capital at the optimal divestment barrier $X^m(k)$ must be equal to its selling price. This means that the following holds

$$W_K\left(X^m(k), k\right) = 1. \qquad (5.12)$$

The optimality condition for $X^m(k)$ requires the high-contact condition:

$$W_{KX}\left(X^m(k), k\right) = 0. \qquad (5.13)$$

When the firm switches from the marginal divestment mode to the firm sale mode we require that the marginal values of capital from the respective policies are equal. Specifically, it must hold that

$$\lim_{k\downarrow K^*} W_K\left(X^m(k),k\right) = \lim_{k\uparrow K^*} W_K\left(X^m(k),k\right). \tag{5.14}$$

If the equality did not hold at K^*, the firm would increase its value by choosing another point to switch from partial to whole-firm divestment. The optimal firm sale is triggered at $X^e(k)$ and the value must satisfy the value matching condition:

$$W\left(X^e(k),k\right) = ak + A. \tag{5.15}$$

The condition means that the firm value must be equal to the proceeds from the sale. The optimality of the endogenous trigger requires that the value function is differentiable at the trigger, which leads to the smooth pasting condition:

$$W_X\left(X^e(k),k\right) = 0. \tag{5.16}$$

Before we characterize the solution of the divestment problem (5.2), let us define

$$R(k) \equiv \left[\gamma\left(a+\frac{A}{k}\right)\right]^{-\beta}\left[(1-\beta)a + \gamma\beta\left(a+\frac{A}{k}\right)\right] - 1. \tag{5.17}$$

Suppose $A > 0$ and $a < a^*$, and let K^* be the unique $k \geq \frac{\gamma A}{1-a\gamma}$ that satisfies $R(k) = 0$. If $a \geq a^*$, let $K^* = \infty$.

Proposition 5.3 *Suppose $A > 0$ and (X_0, K_0) is at or above the relevant triggers characterized below. The optimal divestment policy is characterized by the marginal divestment barrier*

$$X^m(k) = \frac{\beta}{\beta-1}\frac{1}{\gamma}(r-\mu)k^{1-\gamma} \quad \text{if } k > K^*$$

and the firm sale trigger is

$$X^e(k) = \frac{\beta}{\beta-1}(r-\mu)(ak+A)k^{-\gamma} \quad \text{if } k \leq K^*.$$

The firm value is given by

$$W(x,k) = \begin{cases} B_3(k)x^\beta + \frac{1}{r-\mu}xk^\gamma & \text{if } k \geq K^* \text{and } x \geq X^e(k) \\ B_4(k)x^\beta + \frac{1}{r-\mu}xk^\gamma & \text{if } K^e \leq k \leq K^* \text{ and } x \geq X^e(k), \end{cases} \tag{5.18}$$

FIGURE 5.2. Divestment triggers with affine firm-sale premium ($A > 0$) and $a < a^*$. The firm divests partially following the barrier control at $X^m(k)$ as long as $k > K^*$. If $k \leq K^*$ the firm divests the remaining capital at trigger $X^e(k)$.

where

$$B_3(k) = \frac{1}{\beta - 1} \frac{1}{\beta (\gamma - 1) + 1} \left(k X^m(k)^{-\beta} - K^* X^m(K^*)^{-\beta} \right) + B_4(K^*),$$

$$B_4(k) = \left(ak + A - \frac{1}{r - \mu} X^e(k) k^\gamma \right) X^e(k)^{-\beta},$$

and β is as characterized in Proposition 5.1.

5.4   Analysis and implications

Proposition 5.3 characterizes the optimal divestment path. The optimal policy is illustrated in Figure 5.2 and can be described as follows. The firm divests marginally if the capital level is relatively high, above K^*, and whenever x reaches the divestment barrier $X^m(k)$. As soon as capital reaches K^*, the firm stops partial divestment. This is confirmed by Proposition 5.3, which states that partial divestment stops at $X^m(K^*)$ and firm sale is triggered by $X^e(K^*)$. As in general $X^m(K^*)$ will exceed $X^e(K^*)$, the optimal divestment path is characterized by an anticipation region, in which the firm does not divest marginally. Instead it waits until a sufficiently negative profitability shock occurs. This triggers firm sale and exit.

Figure 5.2 clearly illustrates the prediction of the model on the relationship between firm size and divestment policies. Large firms divest partially and small firms divest by firm sale. This prediction finds a strong confirmation in the evidence presented by Maksimovic and Phillips (2001). They find that the average firm that sells partial capital (partial divisions) has revenues of $1.859 billion and operates 23.7 plants, and

the average firm that sells in a merger has revenues of \$51 million and operates 1.78 plants.

An interesting special case is a premium with only the fixed component $A > 0$ and no proportional one, that is $a = 1$. In this case, K^* can be characterized explicitly by

$$K^* = \frac{\gamma A}{1 - \gamma}.$$

The firm size at which the firm is sold is increasing in the fixed premium A and in the level of returns to scale γ. The case of $a = 1$ is also special because the anticipation region $X^m(K^*) - X^e(K^*)$ disappears and the firm continuously moves from partial divestment to full-firm sale.

We are interested in the impact of parameters characterizing the firm and its environment on the choice between partial divestment and firm sale. We first consider the effects of uncertainty represented by the volatility parameter σ in the X_t process.

Proposition 5.4 a^* *decreases in* σ. K^* *increases in* σ *if* $a \in (1, a^*)$.

The proposition states that the effect of uncertainty on the preference between the flexibility of partial divestment and the premium of firm sale is unequivocal. The cutoff level of a that makes the firm to opt for full-firm sale decreases in the level of uncertainty. This means that in a more uncertain environment, the firm prefers full-firm sale for a larger set of parameters. This same kind of prediction is implied by the effect of σ on K^*: the firm exits with higher level of capital after some partial divestment.

These results may seem surprising at first. From the standard real options theory we know that higher uncertainty increases the value of waiting. One might expect that the flexibility advantage of partial divestment is more valuable in a more uncertain market. We find the opposite and the intuition for our result is the following. Firm sale is one irreversible real option and, as such, has a substantial value created by the value of waiting. Partial gradual divestment forms a sequence of real options, and despite the fact that these marginal divestment decisions are irreversible, the whole policy is, in a sense, less irreversible than firm sale. Hence the optimal gradual investment policy takes less into account the value of waiting and the value of the policy will be less responsive to the parameters affecting the value of flexibility.[7] Consequently, the value of firm sale is more responsive to the changes in uncertainty than is the value of gradual partial divestment and the former value increases more in σ making firm sale more attractive.

---

[7] These observations are similar to Malchow-Moeller and Thorsen (2005) who constrast repeated investment options and a single investment option.

Proposition 5.5 a^* *increases in* μ. K^* *decreases in* μ *if* $a \in (1, a^*)$.

The result in the proposition implies that in a more declining market, the option to sell the whole firm and exit becomes more preferable over gradual divestment. In particular, with lower μ, the cutoff premium a^* decreases and the size of full-firm sale K^* increases. Intuitively, in a more declining market, there is less room to benefit from the flexibility of gradual divestment.

5.5   Industry-specific capital and divestment

The price of capital has been fixed in the above formulation. Arguably, in a declining market the selling prices of capital are linked with the state of the market. One reason for prices changing together with market/profitability shocks is industry-specificity of capital. If capital is less productive outside industry, then, after a negative industry-related shock, demand for displaced capital falls and prices decrease. The argument is in line with the industry-equilibrium model of Shleifer and Vishny (1992). Their paper explicitly models potential buyers of displaced capital and predicts that negative industry-specific shocks and financing constraints will result in depressed prices of used capital.

We model these effects in a reduced form by linking the capital price P_t with the market/productivity process X_t. Specifically, we suppose that the evolution of X_t and P_t is given by

$$dX_t = \mu_X X_t dt + \sigma_X X_t (dZ_X)_t$$

and

$$dP_t = \mu_P P_t dt + \sigma_P P_t (dZ_P)_t,$$

where $\mathbb{E}[(dZ_X)_t (dZ_P)_t] = \rho dt$. We interpret the correlation coefficient ρ as the parameter measuring the industry-specificity of capital. A high positive ρ means that capital is industry specific and a decline in X_t results, on average, in a deflated capital price. To ensure that the problem is well defined and has a finite solution we assume that $\mu_X < r$ and $\sigma_X^2 - 2\rho\sigma_X\sigma_P + \sigma_P^2 > 0$.

The extension with variable capital price adds to the complexity of the model. In order to stay in a tractable environment we assume in this section that the whole firm sells only at a proportional premium, that is $A = 0$ and $a \geq 1$. To summarize, a unit of capital divested partially at time t sells at price P_t, and the firm holding K_t units of capital sells at $aP_t K_t$.

In this setup we are interested in the impact of industry-specificity of capital on the optimal divestment policy. We obtain the following result.

Proposition 5.6 *The more industry-specific is capital (the higher is ρ), the more preferable is gradual partial divestment over firm sale.*

The intuition for the result is related to the value of waiting created by the divestment options. The usual prediction of the real options theory is that in an environment as in this section, the value of waiting decreases if productivity and capital price are more correlated (see, e.g., Hartman and Hendrickson (2002)). As discussed in Section 5.4, the value of waiting is larger for the single option to sell the whole firm than for the sequence of marginal options to divest partially. Thus increasing ρ decreases the value of firm sale more than the value of gradual divestment. To put it differently, when capital is highly industry-specific (high ρ), then, after waiting for the market to deteriorate sufficiently to trigger full-firm sale, the firm will, with high probability, sell its capital at low prices. Consequently, the firm's preference moves towards gradual divestment.

5.6   Conclusions

The chapter has studied divestment decisions and addressed directly the trade-off between the flexibility of gradual divestment and the price premium from full-firm sale. It provides analytical results for firm values and optimal divestment policies under alternative premium-discount structures. In particular, if the firm-sale premium is affine, the firm optimally divests marginal units of capital in a declining market until its size reaches a certain threshold. Subsequently, but after an anticipation phase in which the state of market falls to a sufficiently low level, the remaining capital is sold with the whole firm.

The model produces a number of novel predictions on the optimal choice of divestment policy and, specifically, on the choice between partial divestment and firm sale. We analyze the impact of displaced capital discount, firm sale premium, firm size, profit volatility, market growth and industry-specificity of capital. Future empirical research could directly test these predictions.

Future research should also explore if the same mechanisms that are described in this chapter carry over when competition and potential buyers of capital are modeled explicitly. It may be particularly interesting to study a dynamic oligopoly model of a shrinking industry in which firms play a war of attrition as, for example, in Murto (2004), but then to allow firms to undertake partial divestment and takeovers.

The framework presented in the chapter can be adapted to study the other side the capacity adjustment decision, namely investment. It will be interesting to consider a combination of gradual capital expansion and discrete technological change, analo-

gously to capital downsizing and firm sale analyzed in this chapter. The problem of capital accumulation and technology investment has received considerable attention in deterministic models (see, e.g., Feichtinger, Hartl, Kort and Veliov (2006)), but has not been addressed in the stochastic framework of real options.

5.A   Appendix: Proofs

Proof of Proposition 5.1. Solving (5.3) subject to (5.4)-(5.6), we obtain that

$$X^m(k) = \frac{\beta}{\beta - 1} \frac{1}{\gamma} \left(r - \mu \right) k^{1-\gamma},$$

and, if $x \geq X^m(k)$,

$$V^m(x, k) = \frac{1}{1 - \beta} \frac{k}{1 - \beta \left(1 - \gamma \right)} \left(\frac{x}{X^m(k)} \right)^\beta + \frac{1}{r - \mu} x k^\gamma.$$

The solution to (5.7) subject to (5.8)-(5.10) is

$$X^e(k) = \frac{\beta}{\beta - 1} \left(r - \mu \right) \left(a + \frac{A}{k} \right) k^{1-\gamma},$$

and, if $x \geq X^e(k)$, then

$$V^e(x, k) = \left[a - \frac{\gamma \beta}{\beta - 1} \left(a + \frac{A}{k} \right) \right] \left(\frac{x}{X^e(k)} \right)^\beta + \frac{1}{r - \mu} x k^\gamma.$$

Now suppose that $A = 0$ and $x \geq \max \left\{ X^e(k), X^m(k) \right\}$. Using the value functions characterized above, we have that

$$V^m(x, k) - V^e(x, k) = \frac{k}{1 - \beta} \left(\frac{x}{X^e(k)} \right)^\beta \left[\frac{a^\beta \gamma^\beta}{1 - \beta \left(1 - \gamma \right)} - a \right].$$

The sign of the expression depends on the sign of the term in the square brackets. This means that if $a \geq a^*$ then $V^m(x, k) \leq V^e(x, k)$ and if $a < a^*$ then $V^m(x, k) > V^e(x, k)$.

In the case of $a < a^*$, the value of gradual divestment always exceeds the value of firm sale, so it is never optimal for the firm to choose the latter strategy. It follows that the optimal trigger policy of the firm with both divestment strategies available is given by $X^m(k)$ and its value W is equal to the value of the firm with marginal divestment $V^m(x, k)$.

In the case of $a \geq a^*$, the value of strategy comprising of *only* gradual divestment is always below the value of optimal firm sale. To conclude that the firm does not divest

gradually, we still need to rule out a strategy consisting of some gradual divestment followed by firm sale. Suppose the firm divests a marginal unit of capital before the whole firm is sold. The marginal value of capital that is sold optimally by partial divestment is equal to $V_K^m(x, k)$ if $x > X^m(k)$ and equal to 1 if $x \leq X^m(k)$. In the first case, if $x > X^m(k)$, comparing this marginal value with the marginal value of capital from firm sale, we have that

$$V_k^m(x, k) - V_k^e(x, k) = \frac{1}{1 - \beta} \left(\frac{x}{X^e(k)} \right)^\beta \left\{ a^\beta \gamma^\beta - [1 - \beta(1 - \gamma)] a \right\} \leq 0,$$

which is non-positive because $a \geq a^*$. In the second case, if $x = X^m(k)$, the difference in marginal values is

$$1 - V_k^e(X^m(k), k) = \frac{1}{1 - \beta} \left\{ 1 - [1 - \beta(1 - \gamma)] a^{1-\beta} \gamma^{-\beta} \right\} \leq 0.$$

The last inequality holds because $a \geq a^*$. It can be easily verified that for $X^e(k) \leq x \leq X^m(k)$, $V_k^e(X^m, k)$ is decreasing in x, so the difference $1 - V_k^e(x, k)$ remains non-positive (to see that $V_k^e(X^m, k)$ is decreasing in this interval, observe that $V_{xk}^e(X^m(k), k) < 0$ and that $V_{xk}^e(x, k)$ is a convex function on the relevant interval). It follows that the marginal value of capital sold by the firm sale always exceeds the marginal value of capital from partial divestment, so the maximizing firm never chooses to divest partially. ■

Proof of Lemma 5.2. The same steps that in the proof of Proposition 5.1 lead to the following formula for the difference between the values:

$$V^m(x, k) - V^e(x, k) = \frac{k}{1 - \beta} \left(\frac{x}{X^e(k)} \right)^\beta \left[\xi - \frac{A}{k} \right],$$

where $\xi \equiv a^\beta \gamma^\beta [1 - \beta(1 - \gamma)]^{-1} - a$. It was also shown there that $\xi \leq 0$ is equivalent to $a \geq a^*$. It follows that $a \geq a^*$ implies that $\xi \leq A/k$ for all $k \geq 0$. Thus $a \geq a^*$ implies that $V^e(x, k) \geq V^m(x, k)$.

In the case of $a < a^*$, it holds that $\xi > 0$. So there exists $\tilde{k} > 0$ such that $\xi = A/\tilde{k}$. Moreover, $V^m(x, k) > V^e(x, k)$ if $k > \tilde{k}$, and $V^m(x, k) < V^e(x, k)$ if $k < \tilde{k}$. ■

Proof of Proposition 5.3. It is straightforward to verify that (5.18) satisfies (5.11)-(5.13) and (5.15)-(5.16) for a given K^*. Note that $\lim_{k \downarrow K^*} W_K(X^m(k), k) = 1$. Now we consider two cases to verify (5.14). First, if K^* is such that $X^e(K^*) > X^m(K^*)$, then the firm is sold at $X^m(K^*)$, and so $\lim_{k \uparrow K^*} W_K(X^m(k), k) = a$. It follows that, as long as $a > 1$, (5.14) cannot be satisfied if $X^e(K^*) > X^m(K^*)$. Second, we consider

$X^e(K^*) \leq X^m(K^*)$, which can be shown to be equivalent to $k \geq \frac{\gamma A}{1-a\gamma}$. Applying then (5.14) to (5.18) we obtain that K^* must satisfy $R(K^*) = 0$. To verify that K^* is unique in the case of $a < a^*$, we show that there is a unique root to $R(k) = 0$ if $k \geq \frac{\gamma A}{1-a\gamma}$. It can be easily checked that $R'(k) < 0$ if $k > \frac{\gamma A}{1-a\gamma}$. Moreover, $R(\frac{\gamma A}{1-a\gamma}) = (1-\beta)^{-1}(a-1) \geq 0$. So $R(k)$ is monotonically decreasing starting from a positive value. Whether $R(k)$ has a root for $k > \frac{\gamma A}{1-a\gamma}$ depends on a. Note that $\lim_{k \to \infty} R(k) = \gamma^{-\beta} a^{1-\beta}[1 - \beta(1-\gamma)] - 1$ is negative if $a < a^*$ and positive if $a > a^*$. We conclude that if $a < a^*$ there exists a unique finite K^* such that (5.14) holds. If $a \geq a^*$, the marginal value of capital sold with the whole firm always exceeds the marginal value of capital sold partially and $K^* = \infty$. ∎

Proof of Proposition 5.4. We first consider the effect on a^*. σ influences a^* via β. Taking the derivative of a^* with respect to β we have that

$$\frac{da^*}{d\beta} = \frac{a^* \eta_1}{(1-\beta)^2},$$

where

$$\eta_1 = \frac{(1-\gamma)(1-\beta)}{1-\beta(1-\gamma)} - \log\left[\frac{1-\beta(1-\gamma)}{\gamma}\right].$$

The sign of the derivative depends on the sign of η_1, which is a sum of a positive and negative term. We now show that η_1 is always less or equal to zero. Observe that η_1 increases in $\beta \leq 0$:

$$\frac{d\eta_1}{d\beta} = \frac{(1-\gamma)^2(1-\beta)}{[1-\beta(1-\gamma)]^2} \geq 0.$$

Moreover, $\lim_{\beta \to 0} \eta_1 = 1 - \gamma + \log\gamma < 0$ for all $\gamma \in (0,1)$. Thus η_1 is non-positive for all $\beta \leq 0$ and consequently $da^*/d\beta \leq 0$. Finally, it is straightforward to verify that $d\beta/d\sigma > 0$ so $da^*/d\sigma \geq 0$ as stated in the proposition.

Now consider the derivative of K^* with respect to σ. Recall that if $a \in (1, a^*)$, then K^* is the unique $k \geq \gamma A/(1-a\gamma)$ such that $R(k) = 0$. Thus

$$\frac{dK^*}{d\sigma} = -\frac{\partial R/\partial \sigma}{\partial R/\partial K^*}.$$

First, let $\eta_2 = [\gamma(a + A/k)]^{-\beta}$ and consider $\partial R/\partial \beta$:

$$\frac{dR}{d\beta} = \eta_2\left\{-\log\left[\gamma\left(a + \frac{A}{k}\right)\right]\left[(1-\beta)a + \gamma\beta\left(a + \frac{A}{k}\right)\right] - a + \gamma\left(a + \frac{A}{k}\right)\right\}$$

$$= -\frac{1}{\beta}(a\eta_2 - 1 - \log\eta_2) > 0,$$

where in the second equality we twice use substitutions implied by $R(k) = 0$, and the inequality follows from the observation that $\eta_2 - 1 \geq \log \eta_2$ for all positive η_2 with equality holding only at $\eta_2 = 1$. Combined with the previous observation that $d\beta/d\sigma > 0$, we have that $dR/d\sigma > 0$. Second, consider $\partial R/\partial K^*$:

$$\frac{\partial R}{\partial K^*} = -\beta \left(\beta - 1\right) \frac{\gamma A}{k^2} \left[\gamma \left(a + \frac{A}{k}\right)\right]^{-\beta-1} \left[\gamma \left(a + \frac{A}{k}\right) - a\right] < 0.$$

The inequality follows from the fact that $\gamma a \leq \gamma \left(a + A/k\right) \leq 1$ for $k \geq \gamma A/ \left(1 - a\gamma\right)$. Combining the above observations we obtain that $dK^*/d\sigma > 0$. ∎

Proof of Proposition 5.5. The proof is very similar to the proof of Proposition 5.4. μ affects a^* and K^* only via β. The only difference with the effect of σ in Proposition 5.4 is that—as can be readily verified—$d\beta/d\sigma > 0$. Applying this to the derivatives in the proof of Proposition 5.4 we obtain the result. ∎

Proof of Proposition 5.6. The firm optimization problem is now the following

$$W(X_t, P_t, K_t) = \sup_{\tau} \sup_{\{dK_{t+s}\}} \mathbb{E}_t \left[\int_0^{\tau-t} e^{-rs} \pi(X_{t+s}, P_{t+s}, K_{t+s}) ds \right.$$
$$\left. + \int_0^{\tau-t} e^{-rs} P_{t+s} dK_{t+s} + e^{-r(\tau-t)} a P_\tau K_\tau \right]. \quad (5.19)$$

We take the same strategy as in Section 5.3.1 and Proposition 5.1. That is we suppose that (X_0, P_0, K_0) is at or above the relevant triggers and we consider two limit cases, one in which the firm has available only partial divestment and one in which the firm can only divest all capital at once. Both cases are straightforward simplifications of the more general optimization problem (5.19). Denote by $V^m(x, p, k)$ the value function of the firm following optimal partial divestment and by $V^e(x, p, k)$ the value function of the firm following optimal firm-sale policy. The value functions $V^\theta(x, p, k)$, $\theta \in \{m, e\}$, must satisfy the following partial differential equation (where we omit the function arguments for brevity):

$$rV^\theta = \frac{1}{2}\sigma_X^2 x^2 V_{XX}^\theta + \frac{1}{2}\sigma_P^2 p^2 V_{PP}^\theta + \rho\sigma_X\sigma_P xp V_{XP}^\theta + \mu_X x V_X^\theta + \mu_P p V_P^\theta + xk^\gamma. \quad (5.20)$$

Using that $V^\theta(x, p, k)$ is homogeneous of degree one in x and p, we can simplify the problem and reduce one state variable. Let $y = x/p$ and $v^\theta(y, k) = V^\theta(x/p, 1, k) = V^\theta(x, p, k)/p$. This implies that $V_X^\theta = v_Y^\theta$, $V_{XX}^\theta = v_{YY}^\theta/p$, $V_P^\theta = v^\theta - yv_Y^\theta$, $V_{PP}^\theta =$

$y^2 v^{\theta}{}_{YY}/p$ and $V^{\theta}_{XP} = -y v^{\theta}_{YY}/p$. Then we can rewrite (5.20) in terms of v^{θ}:

$$\left(r - \mu_P\right) v^{\theta} = \left(\frac{1}{2}\sigma_X^2 - \rho \sigma_X \sigma_P + \frac{1}{2}\sigma_P^2\right) y^2 v^{\theta}{}_{YY} + \left(\mu_X - \mu_P\right) y v^{\theta}_Y + y k^{\gamma}.$$

The two ordinary differential equations for $\theta = m$ and $\theta = e$ have known general analytical solutions and are solved for the optimal value and divestment policy by setting appropriate boundary conditions. In the case of $\theta = m$, the optimal policy takes the form of barrier control at lower boundary $Y^m(k)$ in the space (y, k). We set the boundary conditions similar to conditions (5.4)-(5.6), i.e. $v_X^m(Y^m(k), k) = 1$, $v_{XK}^m(Y^m(k), k) = 0$ and the finiteness condition as y goes infinity. In the case of $\theta = e$, the optimal policy takes the form of an exit trigger $Y^e(k)$. The boundary conditions in this case are similar to the conditions (5.8)-(5.10), i.e. $v^e(Y^e(k), k) = ak$, $v_X^e(Y^e(k), k) = 0$ and the finiteness condition as y goes infinity.

Applying the boundary conditions we obtain in the case of $\theta = m$ that

$$Y^m(k) = \frac{\beta_1}{\beta_1 - 1} \frac{1}{\gamma} \left(r - \mu_X\right) k^{1-\gamma},$$

and, if $x/p \geq Y^m(k)$,

$$V^m(x, p, k) = p v^m(y, k) = \frac{1}{1 - \beta_1} \frac{pk}{1 - \beta_1 (1 - \gamma)} \left(\frac{x/p}{Y^m(k)}\right)^{\beta} + \frac{1}{r - \mu_X} x k^{\gamma},$$

where β_1 is the negative root of the quadratic equation:

$$\left(\frac{1}{2}\sigma_X^2 - \rho \sigma_X \sigma_P + \frac{1}{2}\sigma_P^2\right) \beta \left(\beta - 1\right) + \left(\mu_X - \mu_P\right) \beta + \mu_P - r = 0. \qquad (5.21)$$

In the case of $\theta = e$, we have

$$Y^e(k) = \frac{\beta_1}{\beta_1 - 1} \left(r - \mu_X\right) a k^{1-\gamma},$$

and, if $x/p \geq Y^e(k)$, then

$$V^e(x, p, k) = p v^e(y, k) = ap \frac{1 - \beta_1 (1 - \gamma)}{1 - \beta_1} \left(\frac{x/p}{Y^e(k)}\right)^{\beta} + \frac{1}{r - \mu_X} x k^{\gamma}.$$

As in Proposition 5.1 we compare the values from the two limit policies, namely V^m and V^e. Straightforward calculations following the argument in Proposition 5.1 lead to the conclusion that there is a threshold level of a^* on a such that partial divestment is preferable over firm sale if $a < a^*$, and if $a \geq a^*$ the firm will optimally sell at once

without partial divestment. It can be verified that

$$a^* = \frac{1}{\gamma} \left[\frac{1 - \beta_1 (1 - \gamma)}{\gamma} \right]^{\frac{1}{\beta_1 - 1}}.$$

The derivative of a^* with respect to β_1 is the same as the one analyzed in the proof of Proposition 5.4, and it was shown there that $da^*/d\beta_1 \leq 0$. Differentiating (5.21) we obtain that $d\beta_1/d\rho < 0$. It follows that $da^*/d\rho \geq 0$, or in words, that with higher ρ the firm requires more premium to optimally choose firm sale over partial divestment. ∎

References

Abel, A. B.: 1983, Optimal investment under uncertainty, *American Economic Review* **73**, 228–233.

Abel, A. and Eberly, J.: 1996, Optimal investment with costly reversibility, *Review of Economic Studies* **63**, 581–593.

Abel, A. and Eberly, J.: 1999, The effects of irreversibility and uncertainty on capital accumulation, *Journal of Monetary Economics* **44**, 339–377.

Abel, A. and Eberly, J.: 2004, Q theory without adjustment costs and cash flow effects without financing constraints. University of Pennsylvania, Working paper.

Acharya, V., Almeida, H. and Campello, M.: 2007, Is cash negative debt? A hedging perspective on corporate financial policies, *Journal of Financial Intermediation* **16**, 515–554.

Acharya, V., Davydenko, S. and Strebulaev, I.: 2007, Cash holdings and credit spreads. Working Paper, London School of Business.

Acharya, V., Huang, J., Subrahmanyam, M. and Sundaram, R.: 2006, When does strategic debt-service matter?, *Economic Theory* **29**, 363–378.

Almeida, H., Campello, M. and Weisbach, M.: 2004, The cash flow sensitivity of cash, *Journal of Finance* **59**, 1777–1804.

Altinkiliç, O. and Hansen, R.: 2000, Are there economies of scale in underwriting fees? Evidence of rising external financing costs, *Review of Financial Studies* **13**, 191–218.

Andrade, G., Mitchell, M. and Stafford, E.: 2001, New evidence and perspectives on mergers, *Journal of Economic Perspectives* **15**, 103–120.

Bansal, R. and Viswanathan, S.: 1993, No arbitrage and arbitrage pricing: A new approach, *Journal of Finance* **48**, 1231–1262.

Bar-Ilan, A. and Strange, W. C.: 1996, Investment lags, *American Economic Review* **86**, 610–622.

Bates, T., Kahle, K. and Stulz, R.: 2008, Why do U.S. firms hold so much more cash than they used to?, *Journal of Finance*, forthcoming.

Bertola, G. and Caballero, R.: 1994, Irreversibility and aggregate investment, *Review of Economic Studies* **61**, 223–246.

Biais, B., Mariotti, T., Plantin, G. and Rochet, J.: 2007, Dynamic security design: Convergence to continuous time and asset pricing implications, *Review of Economic Studies* **74**, 345–390.

Black, F. and Scholes, M.: 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* **81**, 637–654.

Bo, H. and Lensink, R.: 2005, Is the investment-uncertainty relationship nonlinear? An empirical analysis for the Netherlands, *Economica* **72**, 307–331.

Boone, A. and Mulherin, J.: 2007, How are firms sold?, *Journal of Finance* **62**, 847–875.

Brav, A., Graham, J., Harvey, C. and Michaely, R.: 2005, Payout policy in the 21st century, *Journal of Financial Economics* **77**, 483–527.

Broadie, M., Chernov, M. and Sundaresan, S.: 2007, Optimal debt and equity values in the presence of Chapter 7 and Chapter 11, *Journal of Finance* **62**, 1341–1377.

Bulow, J. and Klemperer, P.: 1999, The generalized war of attrition, *American Economic Review* **89**, 175–189.

Caballero, R. J.: 1991, On the sign of the investment-uncertainty relationship, *American Economic Review* **81**, 279–288.

Caballero, R. J. and Pindyck, R. S.: 1996, Uncertainty, investment, and industry evolution, *International Economic Review* **37**, 641–662.

Dangl, T. and Wirl, F.: 2003, Investment under uncertainty: Calculating the value function when the Bellman equation cannot be solved analytically, *Journal of Economic Dynamics and Control* **28**, 1437–1460.

Davydenko, S.: 2007, When do firms default? A study of the default boundary. Working Paper, University of Toronto.

DeAngelo, H. and DeAngelo, L.: 1990, Dividend policy and financial distress: An empirical investigation of troubled NYSE firms, *Journal of Finance* **45**, 1415–1431.

Décamps, J., Mariotti, T., Rochet, J. and Villeneuve, S.: 2007, Issuance costs and stock return volatility. IDEI Working Paper.

Décamps, J., Mariotti, T. and Villeneuve, S.: 2006, Irreversible investment in alternative projects, *Economic Theory* **28**, 425–448.

Décamps, J.-P. and Mariotti, T.: 2004, Investment timing and learning externalities, *Journal of Economic Theory* **118**, 80–102.

Décamps, J. and Villeneuve, S.: 2007, Optimal dividend policy and growth option, *Finance and Stochastics* **11**, 3–27.

DeMarzo, P. and Sannikov, Y.: 2006, Optimal security design and dynamic capital structure in a continuous-time agency model, *Journal of Finance* **61**, 2681–2724.

DeMarzo, P. and Sannikov, Y.: 2007, A learning model of dividend smoothing. Working Paper, Stanford University.

Detemple, J.: 2005, *American-style Derivatives: Valuation and Computation*, Chapman & Hall/CRC, London.

Dixit, A. K. and Pindyck, R. S.: 1994, *Investment under Uncertainty*, Princeton University Press, Princeton, NJ.

Duffie, D. and Lando, D.: 2001, Term structures of credit spreads with incomplete accounting information, *Econometrica* **69**, 633–664.

Dumas, B.: 1991, Super contact and related optimality conditions., *Journal of Economic Dynamics and Control* **15**, 675–685.

Dutta, P. K. and Rustichini, A.: 1993, A theory of stopping time games with applications to product innovations and asset sales, *Economic Theory* **3**, 743–763.

Eom, Y., Helwege, J. and Huang, J.: 2004, Structural models of corporate bond pricing: An empirical analysis, *Review of Financial Studies* **17**, 499–544.

Fan, H. and Sundaresan, S.: 2000, Debt valuation, renegotiation, and optimal dividend policy, *Review of Financial Studies* **13**, 1057–1099.

Faulkender, M. and Wang, R.: 2006, Corporate financial policy and the value of cash, *Journal of Finance* **61**, 1957–1990.

Feichtinger, G., Hartl, R., Kort, P. and Veliov, V.: 2006, Anticipation effects of technological progress on capital accumulation: a vintage capital approach, *Journal of Economic Theory* **126**, 143–164.

Flaherty, T. M.: 1980, Dynamic limit pricing, barriers to entry, and rational firms, *Journal of Economic Theory* **23**, 160–182.

Fudenberg, D. and Tirole, J.: 1986, A theory of exit in duopoly, *Econometrica* **54**, 943–960.

Gamba, A. and Triantis, A.: 2008, The value of financial flexibility, *Journal of Finance*, forthcoming.

Gaskins, D.: 1971, Dynamic limit pricing: Optimal pricing under threat of entry, *Journal of Economic Theory* **3**, 306–322.

Ghemawat, P. and Nalebuff, B.: 1985, Exit, *Rand Journal of Economics* **16**, 184–194.

Ghemawat, P. and Nalebuff, B.: 1990, The devolution of declining industries, *Quarterly Journal of Economics* **105**, 167–186.

Goldstein, R., Ju, N. and Leland, H.: 2001, An EBIT-based model of dynamic capital structure, *Journal of Business* **74**, 483–512.

Gomes, J.: 2001, Financing investment, *American Economic Review* **91**, 1263–1285.

Grenadier, S. R.: 1996, The strategic excercise of options: Development cascades and overbuilding in real estate markets, *Journal of Finance* **51**, 1653–1679.

Grenadier, S. R. and Wang, N.: 2005, Investment timing, agency and information, *Journal of Financial Economics* **75**, 493–533.

Gryglewicz, S., Huisman, K. and Kort, P.: 2008, Finite project life and uncertainty effects on investment, *Journal of Economic Dynamics and Control* **32**, 2191–2213.

Guney, Y., Ozkan, A. and Ozkan, N.: 2007, International evidence on the non-linear impact of leverage on corporate cash holdings, *Journal of Multinational Financial Management* **17**, 45–60.

Guo, X., Miao, J. and Morellec, E.: 2005, Irreversible investment with regime shifts, *Journal of Economic Theory* **122**, 37–59.

Guo, X. and Pham, H.: 2005, Optimal partially reversible investment with entry decision and general production function, *Stochastic Processes and their Applications* **115**, 705–736.

Guo, X. and Shepp, L.: 2001, Some optimal stopping problems with nontrivial boundaries for pricing exotic options, *Journal of Applied Probability* **38**, 647–658.

Hackbarth, D., Hennessy, C. and Leland, H.: 2007, Can the trade-off theory explain debt structure?, *Review of Financial Studies* **20**, 1389–1428.

Hackbarth, D., Miao, J. and Morellec, E.: 2006, Capital structure, credit risk, and macroeconomic conditions, *Journal of Financial Economics* **82**, 519–550.

Han, S. and Qiu, J.: 2007, Corporate precautionary cash holdings, *Journal of Corporate Finance* **13**, 43–57.

Hand, J. and Lev, B.: 2003, *Intangible Assets: Values, Measures, and Risks*, Oxford University Press.

Hartman, R.: 1972, The effects of price and cost uncertainty on investment, *Journal of Economic Theory* **5**, 258–266.

Hartman, R. and Hendrickson, M.: 2002, Optimal partially reversible investment, *Journal of Economic Dynamics and Control* **26**, 483–508.

Hennessy, C. and Whited, T.: 2005, Debt dynamics, *Journal of Finance* **60**, 1129–1165.

Huddart, S., Hughes, J. and Levine, C.: 2001, Public disclosure and dissimulation of insider trades, *Econometrica* **69**, 665–681.

Irvine, P. J. and Pontiff, J.: 2008, Idiosyncratic return volatility, cash flows, and product market competition, *Review of Financial Studies*, forthcoming.

Jeanblanc-Picqué, M. and Shiryaev, A. N.: 1995, Optimization of the flow of dividends, *Russian Mathematical Surveys* **50**, 257–277.

Kamien, M. and Schwartz, N.: 1971, Limit pricing and uncertian entry, *Econometrica* **39**, 441–451.

Kaya, A.: 2007, Repeated signaling games. Working Paper, University of Iowa.

Keppo, J., Moscarini, G., and Smith, L.: 2008, The demand for information: More heat than light, *Journal of Economic Theory* **138**, 21–50.

Leahy, J. V.: 1993, Investment in competitive equilibrum: The optimality of myoptic behavior, *Quarterly Journal of Economics* **108**, 1105–1133.

Leary, M. and Roberts, M.: 2005, Do firms rebalance their capital structures?, *Journal of Finance* **60**, 2575–2619.

Lee, F. C., Wu, C. and Wei, K.: 1990, The heterogeneous investment horizon and the capital asset pricing model: Theory and implications, *Journal of Financial and Quantitative Analysis* **25**, 361–376.

Leland, H. E.: 1994, Corporate debt value, bond covenants, and optimal capital structure, *Journal of Finance* **49**, 1213–1252.

Leland, H. and Toft, K.: 1996, Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads, *Journal of Finance* **51**, 987–1019.

Li, S.: 2003, A valuation model for firms with stochastic earnings, *Applied Mathematical Finance* **10**, 229–243.

Lieberman, M.: 1990, Exit from declining industries: 'shakeout' or 'stakeout'?, *RAND Journal of Economics* **21**, 538–554.

Lins, K., Servaes, H. and Tufano, P.: 2007, What drives corporate liquidity? An international survey of strategic cash and lines of credit. Working Paper, Harvard Business School.

Lintner, J.: 1956, Distribution of incomes of corporations among dividens, retained earnings, and taxes, *American Economic Review* **46**, 97–113.

Liptser, R. S. and Shiryaev, A. N.: 2001, *Statistics of Random Processes: I General Theory*, Springer.

Løkka, A. and Zervos, M.: 2008, Optimal dividend and issuance of equity policies in the presence of proportional costs, *Insurance: Mathematics and Economics*, forthcoming.

Majd, S. and Pindyck, R. S.: 1987, Time to build, option value and investment decisions, *Journal of Financial Economics* **18**, 7–27.

Maksimovic, V. and Phillips, G.: 2001, The market for corporate assets: Who engages in mergers and asset sales and are there efficiency gains?, *Journal of Finance* **56**, 2019–2065.

Malchow-Moeller, N. and Thorsen, B. J.: 2005, Repeated real options: Optimal invest-
   ment behaviour and a good rule of thumb, *Journal of Economic Dynamics and
   Control* **29**, 1025–1041.

Matthews, S. and Mirman, L.: 1983, Equilibrium limit pricing: The effects of private
   information and stochastic demand, *Econometrica* **51**, 981–96.

McDonald, R. and Siegel, D.: 1986, The value of waiting to invest, *Quarterly Journal
   of Economics* **101**, 707–728.

Merton, R.: 1974, On the pricing of corporate debt: The risk structure of interest rates,
   Journal of Finance **29**, 449–470.

Merton, R. C.: 1973, An intertemporal capital asset pricing model, *Econometrica*
   41, 867–887.

Merton, R. C.: 1976, Option pricing when underlying asset returns are discontinuous,
   Journal of Financial Economics **3**, 125–144.

Mester, L. J.: 1992, Perpetual signalling with imperfectly correlated costs, *Rand Jour-
   nal of Economics* **23**, 548–563.

Metcalf, G. E. and Hassett, K. A.: 1995, Investment under alternative return assump-
   tions: Comparing random walks and mean reversion, *Journal of Economic Dy-
   namics and Control* **19**, 1471–1488.

Miao, J.: 2005, Optimal capital structure and industry dynamics, *Journal of Finance*
   60, 2621–2659.

Milgrom, P. and Roberts, J.: 1982, Limit pricing and entry under incomplete informa-
   tion: An equilibrium analysis, *Econometrica* **50**, 443–459.

Miller, M. and Rock, K.: 1985, Dividend policy under asymmetric information, *Journal
   of Finance* **40**, 1031–1051.

Morellec, E. and Zhdanov, A.: 2005, The dynamics of mergers and acquisitions, *Journal
   of Financial Economcs* **77**, 649–672.

Moscarini, G.: 2005, Job matching and the wage distribution, *Econometrica* **73**, 481–
   516.

Murto, P.: 2004, Exit in duopoly under uncertainty, *Rand Journal of Economics*
   35, 111–127.

Opler, T., Pinkowitz, L., Stulz, R. and Williamson, R.: 1999, The determinants and implications of corporate cash holdings, *Journal of Financial Economics* **52**, 3–46.

Pawlina, G. and Kort, P. M.: 2005, Investment under uncertainty and policy change, *Journal of Economic Dynamics and Control* **29**, 1193–1209.

Pennings, E.: 2004, Optimal pricing and quality choice when investment in quality is irreversible, *Journal of Industrial Economics* **52**, 569–589.

Peskir, G.: 1998, Optimal stopping of the maximum process: The maximality principle, *Annals of Probability* **26**, 1614–1640.

Peskir, G. and Shiryaev, A.: 2006, *Optimal Stopping and Free-Boundary Problems*, Birkhauser Verlag.

Pindyck, R. S.: 1988, Irreversible investment, capacity choice, and the value of the firm, *American Economic Review* **79**, 969–985.

Pindyck, R. S.: 2004, Volatility and commodity price dynamics, *Journal of Futures Markets* **24**, 1029–1047.

Pulvino, T.: 1998, Do asset fire sales exist? An empirical investigation of commercial aircraft transactions, *Journal of Finance* **53**, 939–978.

Ramey, V. and Shapiro, M.: 2001, Displaced capital: A study of aerospace plant closings, *Journal of Political Economy* **109**, 958–992.

Saloner, G.: 1984, Dynamic equilibrium limit-pricing in an uncertain environment. Working Paper, Massachusetts Institute of Technology.

Sarkar, S.: 2000, On the investment-uncertainty relationship in a real options model, *Journal of Economic Dynamics and Control* **24**, 219–225.

Sarkar, S.: 2003, The effect of mean reversion on investment under uncertainty, *Journal of Economic Dynamics and Control* **28**, 377–396.

Schwartz, E.: 1997, The stochastic behavior of commodity prices: Implications for valuation and hedging, *Journal of Finance* **52**, 923–973.

Schwert, G.: 2000, Hostility in takeovers: In the eyes of the beholder?, *Journal of Finance* **55**, 2599–2640.

Shleifer, A. and Vishny, R.: 1992, Liquidation values and debt capacity: A market equilibrium approach, *Journal of Finance* **47**, 1343–1366.

Smets, F.: 1991, Exporting versus FDI: The effect of uncertainty, irreversibilities and strategic interactions. Working Paper, Yale University.

Smit, H. and Ankum, L.: 1993, A real options and game-theoretic approach to corporate investment strategy under competition, *Financial Management* **22**, 241–250.

Smit, H. and Trigeorgis, L.: 2004, *Strategic Investment: Real Options and Games*, Princeton University Press.

Sufrin, B. and Jones, A.: 2004, *EC Competition Law*, Oxford University Press.

Sundaresan, S. and Wang, N.: 2007, Investment under uncertainty with strategic debt service, *American Economic Review* **97**, 256–261.

Thijssen, J. J., Huisman, K. J. and Kort, P. M.: 2006, Strategic investment under uncertainty and information spillovers, *Economic Theory* **28**, 399–424.

Titman, S. and Wessels, R.: 1988, The determinants of capital structure choice, *Journal of Finance* **43**, 1–19.

Touzi, N. and Vieille, N.: 2002, Continuous-time Dynkin games with mixed strategies, *SIAM Journal of Control and Optimization* **41**, 1073–1088.

Toxvaerd, F.: 2007, Dynamic limit pricing. Working paper, University of Cambridge.

Wong, K. P.: 2007, The effect of uncertainty on investment timing in a real options model, *Journal of Economic Dynamics and Control* **31**, 2152–2167.

Zeira, J.: 1990, Cost uncertainty and the rate of investment, *Journal of Economic Dynamics and Control* **14**, 53–63.

Samenvatting

Dit proefschrift is een bundeling van vier research papers met als gemeenschappelijk thema de economische effecten van tijd, onzekerheid en informatie.

Hoofdstuk 2 levert een bijdrage aan de dynamische speltheorie en, door toepassing van het algemene model, aan de theorie van de industriële organisatie. Het model introduceert een categorie continue signalling games voor twee spelers, waarbij de ongeïnformeerde speler inzet op een door beide spelers waargenomen diffusieproces. Wij veronderstellen dat de betaling van de inzet afhankelijk is van het type van de geïnformeerde speler zoals de tegenspeler dit zelf heeft waargenomen, en dat de geïnformeerde speler andere types kan nadoen, maar wel tegen een bepaalde 'prijs'. Wij tonen aan dat het signalling game gespeeld wordt zolang de inzet binnen de tweezijdige limiet van de onafhankelijke variabele (de inzet) blijft. In een evenwichtstoestand zal de geïnformeerde speler bij een randomized lower trigger laten zien van welk type hij/zij is. De ongeïnformeerde speler ontdekt welk type de andere speler werkelijk vertegenwoordigt door het minimumproces van de inzet te observeren en speelt om de inzet bij een bovengrens die afneemt met het lopende minimum.

Vervolgens passen we het spel toe op een model van dynamische limietprijzen bij een stochastische vraag en leiden daaruit een aantal gevolgtrekkingen af die onmogelijk zijn met tijdsafhankelijke deterministische modellen. Het model van op onvolledige informatie gebaseerde limietprijzen is voor het eerst gebruikt door Milgrom en Roberts (1982) voor een gevestigd bedrijf dat door middel van zijn prijsstelling duidelijk maakt dat toetreding tot de markt niet winstgevend kan zijn en daarmee potentiële toetreders afschrikt. Wij hebben het algemene model voor signalling games geschikt gemaakt om het probleem van de limietprijzen te bestuderen, en wel door het diffusieproces te interpreteren als de stochastische vraag, de geïnformeerde speler als het geves-

tigde bedrijf en de ongeïnformeerde speler als de toetreder. Eén voordeel van deze dynamische opzet is dat deze een prijsevenwicht genereert en vooral dat dit evenwicht de limietprijs van de gevestigde marktpartijen toont. In een evenwichtstoestand verraadt de gevestigde partij door middel van limietprijzen tot welk type hij behoort door zijn prijzen te verhogen wanneer de markt moeilijker toegankelijk wordt voor nieuwe toetreders. Prijsverhogingen in een krimpende markt kunnen dus duiden op het hanteren van limietprijzen om toetreders te weren. Het model impliceert verder dat, hoewel de vraag als een Markov-variabele is vormgegeven, de beslissing om al dan niet tot de markt toe te treden mede wordt ingegeven door historische continuïteit (path dependence) en dat bij de beoordeling van de winstgevendheid van de toetreding niet alleen naar de huidige marktsituatie wordt gekeken, maar ook naar de historische minimumprijs.

Hoofdstuk 3 bevat een bijdrage aan de corporate-financetheorie. In dit hoofdstuk analyseren we de effecten van financiële problemen op de keuze voor bedrijfsfinanciering en andere financiële beslissingen. Anders dan in de tot nu toe gepubliceerde literatuur hebben wij zowel de liquiditeit op korte termijn als de solvabiliteit op lange termijn bestudeerd. De opname van liquiditeitskwesties in het 'contingent claims trade-off model' heeft allerlei implicaties voor corporate finance. Wij tonen aan dat er een belangrijke wisselwerking bestaat tussen liquiditeit en solvabiliteit. Omdat bedrijven met een geringere solvabiliteit minder geldmiddelen nodig hebben om liquiditeitsschokken op te vangen alvorens in staat van insolventie te geraken, leidt een geringere solvabiliteit tot grotere liquiditeit bij bedrijven. Aan de andere kant is de liquiditeit zowel direct van invloed op de financiële beslissingen — het is namelijk duur om geld te lenen om aan de liquiditeitseisen te voldoen — als indirect, via de keuze voor de optimale vermogensstructuur en daardoor op de solvabiliteit van het bedrijf. Het model geeft een reden voor de soms omvangrijke liquide middelen die bedrijven aanhouden en voorziet in een dynamisch liquiditeitsbeleid in lijn met empirische regelmatigheden. De wisselwerking tussen liquiditeit en solvabiliteit maakt dat positieve schokken in de kasstroom worden ingehouden en negatieve schokken leiden tot een daling van de optimale omvang van de liquide middelen. Het gevolg is dat de optimale dividenduitkeringen worden afgevlakt ten opzichte van de kasstromen. Door liquiditeit in het model te introduceren, kunnen we de kritiek op de voorspellende waarde van structurele modellen deels ondervangen. Ten eerste blijkt het uit empirisch onderzoek van Eom, Helwege en Huang (2004) een veel voorkomend probleem van structurele modellen te zijn dat de voorspelde spreiding van de credit spreads te groot is. Ons model voorspelt een geringere spreiding van de credit spreads over bedrijven dan het model zonder liquiditeit. Ten tweede pakken bij het gebruik van standaard structurele modellen de voorspelde leverage ratios vaak te

hoog uit. Door de liquiditeit mee te rekenen, valt het aandeel vreemd vermogen in de waarde van een bedrijf in onze analyse aanzienlijk lager uit.

In hoofdstuk 4 en 5 bestuderen we investeringsbeslissingen van bedrijven. Hoofdstuk 4 gaat nogmaals in op de belangrijke resultaten van de 'reële-optiebenadering' van investeringen, die inhoudt dat het bij toenemende onzekerheid steeds gunstiger wordt om te wachten met investeren, met als gevolg dalende investeringen. In de literatuur over dit onderwerp wordt er standaard van uitgegaan dat de looptijd van investeringen oneindig is. In de huidige economie verandert de technologie echter voortdurend en dat betekent dat bedrijven bij technologische investeringen doorgaans uitgaan van een beperkte levensduur. In strijd met de bestaande theorie blijkt uit de resultaten eveneens dat toenemende onzekerheid juist leidt tot meer investeringen. Dat blijkt vooral het geval te zijn bij geringe onzekerheid en een korte projectduur.

Hoofdstuk 5 bestudeert het optimale desinvesteringsbeleid voor bedrijven: is dat geleidelijk afbouwen en verkopen of het gehele bedrijf ineens van de hand doen? Geleidelijke desinvesteringen bieden grotere flexibiliteit, terwijl het hele bedrijf tegelijk van de hand doen een hogere prijs oplevert. Wij tonen aan dat een groot bedrijf er het beste aan doet eerst enkele bedrijfsonderdelen af te stoten alvorens de rest van het bedrijf ineens te verkopen. Hierbij geldt wel de voorwaarde dat de (hogere) prijs uit zowel een vaste als een procentuele component bestaat. In de volgende gevallen blijkt volledige verkoop te verkiezen boven gedeeltelijke verkoop: bij relatief sterk fluctuerende winsten, bij dalende markten en als kapitaal minder sectorspecifiek is.

Ook op methodologisch gebied draagt dit proefschrift bij aan de literatuur door op innovatieve manieren stochastische controletechnieken in te zetten bij de oplossing van nieuwe economische problemen. Hoofdstuk 2 bestudeert met behulp van de theorie van de optimale controle van extremumprocessen hoe we kunnen achterhalen van welk type onbekende, andere spelers zijn. Om te beginnen formuleren we het probleem met twee onveranderlijke Markov-variabelen, namelijk de beloning van het spel en het Bayesiaanse geloof omtrent het type speler dat de ander vertegenwoordigt. Vervolgens laten wij zien dat het oorspronkelijke probleem met zijn complexe Bayesiaanse updating in een veel eenvoudiger probleem kan worden vertaald, waarin de onveranderlijke geloofs-variabele wordt vervangen door het minimumproces van de opbrengstvariabele. Het probleem van de ongeïnformeerde speler kunnen we vervolgens oplossen met behulp van 'optimal stopping of maximum processes' (zie Peskir (1998) en Peskir en Siryaev (2006)) dat daartoe een uiterst flexibel kader biedt.

In hoofdstuk 3 introduceren we onbekende drift- en filterparameters in een model voor twee vormen van onzekerheid: korte liquiditeitsschokken en langdurige onzeker-heid over de solvabiliteit. De kortdurende onzekerheid wordt direct weergegeven door onvoorspelbare Brownse toenamen van de kasstromen. Voor de langdurige onzeker-

heid gaan we ervan uit dat de waarde van de gemiddelde momentane kasstroom aanvankelijk onzeker is maar een bekende distributie kent. De uitkomsten van het stochastische proces worden gebruikt om inzicht te krijgen in de ware aard van het kasstroomproces. Deze werkwijze heeft als aantrekkelijk punt dat aanhoudende liquiditeitsschokken daadwerkelijk worden vertaald in solvabiliteitsschokken (zo duiden aanhoudende negatieve liquiditeitsschokken op geringe winstgevendheid). Door aan het model voor kasstroomontwikkeling een filter toe te voegen, krijgen we een corporate financemodel dat sober en analytisch wendbaar is en tegelijk een grote reikwijdte en voorspellende kracht heeft.

In hoofdstuk 5 hanteren we een combinatie van 'barrier control' en 'optimal stopping' om de kosten en baten af te wegen van geleidelijke of abrupte aanpassingen van het bedrijfsvermogen. Geleidelijke aanpassingen, die als 'barrier control'-probleem in het model zijn opgenomen, bieden het bedrijf meer flexibiliteit. Die flexibiliteit is van waarde in stochastische omgevingen en blijft waardevol, ook als de aanpassing onomkeerbaar is. Daar staat tegenover dat onomkeerbare, abrupte (des)investeringen, in het model weergegeven als stopprobleem, minder flexibiliteit bieden maar vaak een hogere prijs opleveren. De combinatie van 'barrier control' en 'optimal stopping control' is nieuw en in de reële-optiebenadering nog niet eerder toegepast.