Introduction to corporate finance theory

William Mann
Goizueta Business School, Emory University

September 3, 2026

This is the text for a doctoral course in corporate finance theory at Emory University. The course was first taught in fall 2022 without a text. After that semester, I typed up the notes I had used for each lecture into this document, and used it as the course text for the second iteration of the course in spring 2025. The third iteration is expected to be taught in the 2027-2028 academic year, and I will make intermittent revisions until then. As of this writing, this document is still a work in progress and is rough or incomplete in several areas. Please inform me about any errors or suggestions to improve it!

This course is meant to be covered in half a semester, while assuming as little as possible about the student's background with economic theory or mathematics. Therefore, the focus is on deriving the classic qualitative results that are most often cited to ground reduced-form empirical work in corporate finance. At the end of each chapter are past questions that have been used on final exams or qualifying exams for the Emory Finance PhD students.

Each chapter focuses on a certain theme in the literature, and highlights the main results of a few of the best-known papers within that theme. Each chapter also introduces a few technical concepts from game theory or equilibrium theory that are particularly important for the topic at hand. The material in the text is not comprehensive. Instead, it is mainly meant as an introduction to the topics and papers that we will cover in class, allowing that discussion to go deeper and make more efficient use of time.

Many of the models are simplified compared to their original versions. In particular, I do not assume familiarity with dynamic methods nor differential equations. For this reason, all models presented are static, and we only touch on connections between corporate finance theory and options pricing. I also frequently use different notation or labeling of results compared to the original source papers, in order to harmonize the discussion across models.

Contents

Chapter 1

Benchmark models

1.1 Overview

In most of this course, we will study models that include one or more frictions and market failures, so that equilibrium outcomes are inefficient. This is because efficient outcomes don't give us very much to talk about as researchers!

However, it is first important to thoroughly understand the frictionless, efficient case as a benchmark, because any other situation is best seen as a departure from this case. From this understanding we can avoid many fallacies that are common in practice, and focus our attention on the economic mechanisms that are actually important. We will also begin to understand the challenges in proving whether frictions are indeed present, or how big they are.

1.2 Investment

In principle, we have a good answer as to which projects *should* get funded: the answer is, those projects with positive NPV. The central question of an investment model is which projects actually *do* get funded. In what way does the answer differ from the positive-NPV benchmark? Is there too little investment, or too much, or something more complicated happening? Is there anything we can do to improve the situation?

The basic challenge here is obvious: NPV makes sense conceptually, but cannot be measured directly, at least not by the econometrician. Hence the goal of this literature is to understand what other patterns we might look for to know whether equilibrium outcomes are not efficient.

To investigate this, we write down a standard model of the firm's optimal investment decision. This model will lead us to the celebrated "q theory of investment," a straightforward linear model that is testable by regression.

The firm begins the model with a capital stock K. It then chooses an investment amount I to arrive at a new capital stock K'. The cost of this

investment is $\psi(I, K)$. Then profits will be realized according to $V(A, K')$ where $A > 0$ is a productivity parameter. Then the model ends.[1]

To understand the firm's optimal decisions, we start by looking at the first-order condition for the investment decision:

$$\psi_1(I, K) \quad = \quad V'(K')$$

We next adopt a specific parameterization $\psi(I, K) = \frac{1}{2}a \times \frac{I^2}{K}$. This leads to

$$\frac{I}{K} \quad = \quad \frac{1}{a} \times V'(K)$$

Notice the intuition here: The investment *rate* (new investment as a fraction of old capital) is guided by the marginal value of the firm's capital stock.

An obvious way to test this model would be to see if firms indeed invest at higher rates when they have higher marginal values of capital. Unfortunately, as the econometrician, we cannot directly measure $V'(K)$.

However, there is one special case in which it is observable. Suppose the firm's profit function is linear in capital, $V(K') = AK'$ where $A > 0$ is a productivity parameter. In this case we have $V'(K') = V/K'$. Note that both V and K' are observable in the data, at least approximately: V is the firm's enterprise value (the total market value of all its financial liabilities, primarily equity and debt securities). K' is a measure of the firm's capital stock, usually interpreted as property, plant, and equipment on the balance sheet (PP&E), but sometimes interpreted more broadly.

Thus we arrive at a prediction that can actually be tested: With the above assumptions, the firm's investment rate I/K ought to be a linear function of the valuation ratio V/K'. This is known as the q theory of investment. If we find a strong linear relationship between these two, that could be interpreted as evidence that managers are making fairly efficient investment decisions. The coefficient from the regression would be an estimate of $1/a$, so a larger coefficient means there is less convexity in the cost of investing ψ, allowing the firm to change its policies more aggressively.

The intuition for the above reasoning is simple and yet quite deep. As the econometrician, we have no straightforward way to judge which investments were more or less valuable in past data. But we can reasonably expect that market participants made the best possible judgment about this, given the available information. We should expect that the firms with the most valuable projects (in the model, the highest A) both invested at the highest rate, and attracted the highest valuation from investors. If this connection turns out to hold in the data, we could interpret it as evidence that investment choices are being made

[1] It may seem odd for the model to end abruptly after one date. This is an example of a *static* model. One could extend this setup to a *dynamic* model, with the firm continuing to act at all future dates. This is clearly more realistic (and an example is presented at the end of this chapter). However, the one-period model, while it may feel odd, is able to demonstrate the economic ideas that we need to take away. This is often true in finance theory. In fact, most of the standard models in finance were originally developed as static models, and only later extended to dynamic versions. We will focus on static models throughout this course.

about as efficiently as we could hope for. On the other hand, if there is not a strong relationship between valuations and investments, then one possible interpretation is that managers are systematically making bad investment decisions (although as we will see, there are also many other possible interpretations).

V/K' could be seen as a standard accounting multiple (enterprise value to tangible assets), though it is not necessarily one that analysts focus on very heavily in practice. In academic research, V/K' has the special name of Tobin's average q, or just sometimes Tobin's q. The quantity $V'(K')$ is known as marginal q. The important feature of this theory is that by assuming a linear profit function, we conclude that marginal q is equal to average q. This gives us a way to measure marginal q (the important quantity in the first-order condition above) by simply constructing average q in accounting data.[2]

Another important piece of the model, though less obvious on first reading, is the investment cost function $\frac{1}{2}a \times \frac{I^2}{K}$. Intuitively, the most important thing about this function is the convexity in I. If not for this feature, the model would exhibit a corner solution: Each firm would jump immediately at each date to its ideal capital stock, and these capital stocks would be very volatile. This seems counterfactual. Instead, firms are typically investing over sustained periods of time, building up their capital towards some target. The above adjustment cost function will generate this slow-moving behavior. (This is easier to appreciate in the dynamic model at the end of this chapter.) The specification of ψ also builds in the reasonable idea that investment is less difficult as the firm is bigger (larger K). However, it is still clearly ad hoc to some extent, and an ongoing topic in the literature is to consider the effect of other specifications.

The standard framework for the q theory of investment is much more general than the above model. In particular it is dynamic, and allows for the firm to be uncertain about the value of A when it makes its investment decisions. We will not cover the dynamic model in detail, but an example is presented for reference at the end of this chapter. Strikingly, neither of the changes mentioned above affect the form of the regression that we derived. What is still critical, however, is the assumption that profits AK' are linear in the capital stock. If we relax this, the model still predicts a connection between q and investment rate, but it inherits the nonlinearity of the profit function, so that a regression will exhibit a positive coefficient but will not fully capture the relationship.

The above analysis suggests to evaluate the model by running the regression

$$\frac{I}{K} \quad = \quad \alpha \quad + \quad \beta \times \frac{V}{K'} \quad + \quad \varepsilon \tag{1.1}$$

where V/K' is average q as defined above, and I/K is the investment rate.

The exact timing of I, K, and V may vary in practice from one author to another, depending on how the model was set up or what interpretation is desired from the regression. (For example it is common to measure them all at the same time, or to measure I one period ahead of all the others, neither of

[2]The use of the letter q dates back to some much older papers by Tobin and Kaldor, but these historical origins are not important for our purposes.

which is strictly consistent with the model we wrote above.) However, timing details tend to make little difference to the results, given the slow-moving nature of all three variables. So think of the model as motivating a general approach to modeling investment, by regressing investment ratios on valuation ratios, without strictly dictating the timing or other details of the approach.

These regressions may be run at the aggregate level, or at the firm level, with or without firm or time fixed effects. Arguably the model is best suited for aggregate analysis, where the assumption of linear profits seems more plausible.

After running such a regression, researchers often do the following:

- Look at the R^2 to understand how much of investment activity is or isn't explained by average Q.
- Check whether the regression coefficient β seems economically "large" or statistically significant.
- Add other controls such as cash flows to see if they also have explanatory power for investment (according to the theory, they should not).

The traditional findings are disappointing: R^2 is very low (well below 50% in the aggregate regression, and close to zero in firm-level regressions). β seems quite small (on the order of 0.01 to 0.02). Other control variables, especially the firm's cash flow, seem to have more power than Q to explain its investment.

All of this evidence has led to a common view that there are *very* serious frictions and market failures affecting investment in practice, and this is the reason for the apparently poor performance of the q theory of investment. This in turn has motivated other models of investment, such as the ones we will see in future chapters based on screening, signaling, hidden actions, and more.

In particular, the strong predictive power of cash flow for investment has led many researchers to conclude that managers simply invest when they have the money, not so much based on the value of their potential investments. This could be seen as the result of a mistake or misbehavior by managers, or as the result of strong financial constraints such that firms cannot invest without cash flow. A vast empirical literature attempts to disentangle these possibilities.

But before moving on, be aware that there is still considerable debate about what exactly this evidence really means. None of the above diagnostics are strictly speaking "tests" of the theory:

- R^2 will be less than 100% even if the model is correct. We have to make an educated guess how big or small we should have expected it to be.
- Likewise, β can be arbitrarily small even if the model is "correct," because the model tells us that $\beta = 1/a$. If it's extremely costly to adjust investments (in the sense of convexity), then the size of those adjustments will be small in the data, and we will not be able to detect easily that managers are following the model's recommendations even if they try to.
- Most importantly, we never really believed that average q was *exactly* equal to marginal q (i.e. that profits are exactly linear in the capital stock).

Clearly, there are features missing from the model that might break this relationship. Just a few examples would be decreasing returns to scale, market power, and error in measuring the capital stock (for example when firms make large intangible investments). That opens up room for other variables like cash flow to explain investment ratios *statistically* (if they have the right correlation with the gap between average and marginal q), while not necessarily implying that those variables affect investment *causally* (which would be a true violation of the theory). Similar issues can arise if we are very wrong about our specification of ψ, for example if investment features large fixed costs.

Some extremely well-known papers within this enormous literature are Erickson and Whited (2000), Gomes (2001), and Cooper and Ejarque (2003).

The overall point is this: Although the frictionless q theory of investment does not work perfectly in the data, this does not automatically mean that the models we will study in future chapters are any more accurate as a description of reality. That is an empirical issue and heavily debated. The point of this course is simply to lay out the most common models, without taking a strong stand on which ones are closest to being true or false.

1.3 Capital structure

Capital structure is the decision of which financial contracts to use to fund a firm's operations. Understanding the firm's capital structure decision is one of the central questions in finance.

While many different financial contracts are used in practice, the most important distinction is between those that are basically "debt" or "equity." Roughly speaking, debt is a senior obligation to pay a fixed amount of money, while equity is a claim to a fraction of the firm's value left over after satisfying all such obligations. So our investigation of capital structure theory will mostly analyze the firm's choice between these two specific contracts.

1.3.1 Modigliani and Miller (1958)

Any study of this topic begins with Modigliani and Miller (1958), which investigates a frictionless model, and shows that it's suprisingly difficult to see why the decision should matter at all.

More precisely, the model shows that the most salient differences between debt and equity – the higher average rate of return, and higher risk, of an equity contract compared to a debt contract – are *not* logical reasons to prefer one combination of debt or equity finance over another, because these patterns arise even in a model where, by design, the choice does not matter.

Although simple, this result is the most important idea in corporate finance theory, and we could spend many weeks exploring all the ramifications. When you teach finance, you will find that it resurfaces everywhere, and is tremendously helpful in answering questions that seem difficult at first.

We imagine that a firm is thinking about adjusting its capital structure, by issuing some amount of debt to buy back outstanding equity, or vice versa, without changing anything about its real operations. Let D represent the market value of the firm's debt, and let E the market value of the firm's equity, when it arrives at its chosen capital structure. Let V represent the hypothetical value of the firm's equity if it had no debt at all. (We could think of a counterfactual world where it never issued debt in the first place, or we could imagine that it issues enough equity today to buy back all its bonds and prepay all its loans.) We call V the "unlevered equity value" and E the "levered equity value."

Finally, we make two key assumptions about this environment. As with all economic models, the point here is not to make an accurate description of reality, but rather to intentionally shut down many factors (even if they might be important in reality) to see how important the remaining ones are.

- The first key assumption is that the firm's choice of capital structure does not directly influence the cash flows generated by the firm's operations.
- The second key assumption is that there is "no arbitrage," meaning that the prices of financial securities always reflect the present value of the future cash flows that they promise.

We call this environment "perfect capital markets." Then,

Proposition 1.1 (MM Prop 1)**.** *Under perfect capital markets, $V = D + E$.*

That is, no matter what capital structure the firm might choose (no matter the value of D), the firm's total enterprise value will always add up to the value the firm would have with no debt at all.

The original paper gave an argument by contradiction based on arbitrage, but we can be much more direct. Indeed, when you think about the extensive assumptions we are making here, the proposition should seem tautological. We directly assumed that the prices of securities always reflect the present value of the cash flows they promise. Since debt and equity together are the only claims on the firm's cash flows, this means that their combined value must always be equal to the present value of those cash flows. Since we further assumed that those cash flows will not be affected by the relative amounts of debt and equity, their present value must always be V.

This result is often described as saying that the value of a pie (the firm's cash flows) is not affected by how you divide that pie up into slices (tranches of debt and equity, or indeed other securities). Tirole (2006, p.78) illustrates with a special case: Suppose a firm has debt with face value F, and the firm's cash flows will generate some uncertain value R tomorrow. The value of its debt can be described as $D = \min(F, R)$. The value of its equity is $E = \max(0, R - F)$. Then we have $D + E = R$ regardless of the value of F.

Again, none of this should seem surprising when you think about it. The surprising thing is how many *other* things can happen in the setting of perfect capital markets, that are often perceived as being important considerations in

the decision between debt and equity financing, but clearly cannot be given the result above. A key example is given in the paper's Proposition 2:

Proposition 1.2 (MM Proposition 2)**.** *Let D', E' and V' represent the values of D, E, and V at some future date. Let r_D, r_E, and r_V represent the percentage growth of each variable between these two dates. In perfect capital markets,*

$$r_V = \frac{D}{D+E} \times r_D + \frac{E}{D+E} \times r_E \tag{1.2}$$

From this,

$$r_E = r_V + \frac{D}{E}(r_V - r_D) \tag{1.3}$$

In words, equation (1.2) says that the return on the unlevered firm must equal the weighted average return on each of the firm's securities, where the weights are equal to the total value of each category of security. Equation (1.3) then shows that, as the firm's capital structure shifts towards debt, the average rate of return on equity will grow according to a relationship that we can make quite precise: It will equal the rate of return on the unlevered firm r_V, adjusted upwards by a factor that depends on the leverage ratio D/E and the gap $r_V - r_D$ between the unlevered rate of return and the debt rate of return.[3,4]

This result can help address many common fallacies. Most importantly, we have replicated the empirical fact that a firm's equity securities have a higher rate of return in equilibrium than its debt securities. By no-arbitrage reasoning, this carries over to the firm's cost of capital for any *new* issuance as well: Any new equity must be priced to offer investors a higher rate of return than what would be necessary for new debt.

The critical point is that, nevertheless, the firm has no logical reason to favor the "cheap" debt financing over the "expensive" equity financing. We know this because we generated the pattern in an environment where (by Proposition 1) we already know the choice of capital structure does not matter at all. The intuition is that if the firm did use "cheap" debt financing, it would just pile more risk onto the equity investors, who apply a higher discount rate and lower their valuation, leaving the firm's enterprise value unchanged.

[3]The derivations are simple but maybe worth writing out: For (1.2),

$$\begin{aligned} r_V \equiv \frac{V'}{V} - 1 &= \frac{D'+E'}{D+E} - 1 \\ &= \frac{D}{D+E} \times \left(\frac{D'}{D} - 1\right) + \frac{E}{D+E} \times \left(\frac{E'}{E} - 1\right) \\ &= \frac{D}{D+E} \times r_D + \frac{E}{D+E} \times r_E \end{aligned}$$

And (1.3) follows simply by solving for r_E.

[4]A note about terminology: In valuation textbtooks, r_V is typically labeled r_U and is called the "unlevered cost of capital." The term "weighted average cost of capital" (WACC) by convention refers to the *tax-adjusted* average $\frac{D}{D+E} \times r_D \times (1-\tau) + \frac{E}{D+E} \times r_E$. We will ignore taxes in this class, but any valuation textbook would cover this topic in detail.

Thus, in perfect capital markets the firm's choice of liabilities to finance itself is irrelevant. Of course, we do not have to believe this is actually true about the real world. However, in order to claim that the choice *does* matter, we must go back to the model and identify what features we have left out that might change this conclusion. This means going back to the two assumptions that described "perfect capital markets" in Proposition 1.1.

The second of the two assumptions was that the prices of the firm's securities accurately reflect the present value of their promised future cash flows. So you could overturn the irrelevance of capital structure by assuming that there are mispricings and the firm can exploit them. Anecdotally, there do seem to be examples of companies that successfully time the market by issuing equity at overheated valuations, and this idea is formalized in the market-timing theory of Baker and Wurgler (2002). More important for our course is the signaling-based theory of Myers and Majluf (1984), which derives mispricing within the model and generates a unique theory of capital structure that we will cover in detail in Chapter 3.

The other assumption of perfect capital markets was that the choice between debt and equity does not affect the overall cash flows that the firm will generate for its investors. Thus, we might also change our conclusion if there is indeed some mechanism by which the firm's capital structure affects its cash flows. This has been the bigger focus of capital structure research. We will investigate many potential mechanisms connecting capital structure and cash flows, but it turns out that the biggest and most obvious such mechanism is simply the effect of corporate income taxation, as we describe in the next section. This effect in turn is the basis of the so-called "tradeoff theory" of capital structure, which is our final topic for this chapter.

1.3.2 Debt and taxes

In most economies including the US, corporate income is taxable, but interest payments on debt are deductible out of taxes, while dividend payments are not. This creates a very strong incentive towards using debt at the margin, simply to minimize the firm's tax bill. In this sense, by using more debt the firm can actually increase the cash flows it generates, a violation of the assumption of perfect capital markets.[5]

But if the only problem with perfect capital markets was the failure to account for taxes, then we would arrive at another extreme prediction that is again at odds with reality. In this scenario firms should use lots of debt financing, up to the point that interest payments offset *all* taxable income. In reality, firms don't even come close to this behavior. Instead, they exhibit much lower debt

[5]Modigliani and Miller (1958) already acknowledged this story, but their discussion underestimated how important tax effects could be quantitatively. They issued a correction in Modigliani and Miller (1963) that acknowledged and emphasized the "tax shields" of debt more heavily, and suggested that this might be the primary driver of capital structure in practice. Standard valuation methods like WACC and APV are careful to take these tax effects into account, and thus implicitly recommend 100% debt financing for every project.

levels that seem to leave significant tax savings on the table. There are many, many papers on this fact, but a well-known reference is Graham (2000).

So, what stops firms from using more debt? Miller (1977) took up this question in his 1976 presidential address to the American Finance Association. A large part of his discussion focuses on the complicating effect of personal taxes, which has not been a major focus of subsequent literature and which you do not need to study (though it might be worth reading if and when you ever teach corporate finance or valuation). The more important legacy of this paper is that it gives an illuminating early discussion of our next topic, the "tradeoff" theory of capital structure.

1.3.3 Tradeoff theory

The most obvious disincentive to use debt financing, despite the tax benefits it provides, is potential "distress costs" that are incurred when the firm is unable to make the payments on the debt. This is the type of explanation that most people instinctively use to explain why firms do not use more debt in practice. A large debt load feels dangerous, since it clearly increases the probability of being unable to pay that debt.

We can summarize this idea formally as the "tradeoff theory" of capital structure: Assume that there are deadweight costs of defaulting on debt that increase in the firm's overall debt load. Then, at some level of debt the marginal tax benefit is not worth the marginal expected distress costs. The firm considers the "tradeoff" between these two forces and settles down at the optimum.

While this idea may seem natural, let's think carefully about the supposed distress costs that drive the story. First of all, they must represent some kind of economic value that is *destroyed* by the very act of defaulting on debt. That is, distress costs do *not* refer to any value that is simply transferred from the equityholders of the firm to its debtholders, but rather must reflect an amount by which the entire "pie" gets smaller. Second, these costs must increase as a *convex* function of the firm's debt load, in order for there to be an optimal amount of debt between zero and 100% of the firm's financing.

The problem is, it's hard to find any clear evidence of deadweight costs of defaulting on debt that fit this description and are large enough to explain firm's avoidance of debt. Studies typically find that the easily-measured costs, such as legal fees, are tiny compared to the tax savings that firms seem to be giving up. Miller (1977) summarized this situation with the colorful metaphor of "horse and rabbit stew." Thus, much of the research in capital structure that we will be cover can be framed as the search for other, less-obvious costs of using debt, sometimes labeled *indirect* costs, that may suggest a larger "rabbit."

With a little creativity, we can see many possible ideas. For example, at a highly-indebted company, a small run of bad luck could cause employees and investors to lose their incentive to add value to the company, knowing that any such value will simply accrue to the senior lenders. This effect is known as *debt overhang* and we will formalize it in Chapter 4. The firm might rationally avoid this situation by not building up much debt in the first place.

1.3.4 Static and dynamic reasoning

When we look for "indirect costs of debt" in the manner described above, we are implicitly accepting the paradigm of tradeoff theory, and just trying to measure more precisely the benefits and costs that it describes. However, there have also been deeper challenges to the tradeoff paradigm itself, based on several patterns in the data that appear puzzling.

For example, if firms were really balancing tax benefits of debt against costs of defaulting on debt, then we might expect the most aggressive debt usage to be at firms that have the lowest probability of default, so those costs are less salient. Instead, we find the opposite: Large firms with plenty of cash flow use debt, if anything, more conservatively than other firms. This has been highlighted as a puzzle by Graham (2000) and Myers (1993).

Also, we might expect that firms aim to maintain target leverage ratios, which would involve actively issuing more debt when their market values increase. Instead, Baker and Wurgler (2002) note that firms seem to issue equity when valuations are high and never issue more debt in response, leading to long-lasting effects of temporary equity fluctuations.

Based on these patterns, some authors reject the idea that managers act at all in accordance with tradeoff theory, and propose various other explanations for how capital structure is chosen (including the possibility that it is somewhat random and simply not important to the firm).

But it is not so easy to reject tradeoff theory in this manner. As pointed out by Hennessy and Whited (2005), the arguments above are based on interpreting the data through capital structure models that are simplified, and in particular, are *static*. Static and dynamic models often generate similar qualitative patterns (as for example with the q theory of investment earlier), but not always. Hennessy and Whited (2005) show carefully that a dynamic model of capital structure can match the empirical patterns that researchers have used to attack the tradeoff theory. Aside from its contribution to capital structure research, this paper highlights that while static models (the focus of our course) are a good way to start thinking about a topic, we cannot be completely certain that their predictions carry over to dynamic settings until we check.

1.3.5 How do managers actually think?

Even defenders of the tradeoff theory must acknowledge that financial managers themselves typically do not think in the terms described by the theory, at least not explicitly. They incorporate many factors into their decisions of which security to issue and when, and the economic concept of "tax benefits of debt" do not frequently appear on that list. Instead, decisions are driven by their perceptions of market conditions and general appetite for one type of issue versus another, depending on the firm's intentions and the context in which the funds will be used. Is this a clear problem with the model?

Not necessarily. The economic analysis from earlier in this chapter highlighted that an economic incentive exists for debt. Those incentives will assert

themselves even if managers do not explicitly think about taking advantage of them. For example, capital budgeting exercises to guide investment decisions will correctly forecast taxes on the income generated by that investment, and they will correctly deduct interest payments from taxable income. Whether or not the manager is thinking about tax shields of debt strategically, she will find that projects with more debt appear more economically attractive.

This is just one example of a more general principle that economic actors do not necessarily need to be thinking consciously about the forces in our models, in order for those forces to be relevant to their decisions. Certainly an alignment with practitioner thinking makes a model more convincing, but it is not essential, provided we are careful about how we interpret that model.

1.4 Conclusion

We have highlighted the behavior of frictionless models of investment and capital structure models. In the rest of the course, we will consider how various frictions, based on problems of information and incentives, might affect the firm's decisions regarding investment and capital structure. We will be especially interested in knowing whether investment rates are higher or lower than efficient levels, whether there is a clear *disadvantage* of debt financing (that might help to rescue the tradeoff theory), or whether the market potentially misvalues firms' securities.

In general, let's be realistic about where this will lead us: The end result will simply be a list of factors that could plausibly matter for the firm's decisions, in one setting or another. We cannot expect to arrive at a "true" model of investment or capital structure. Indeed, the goal of economic theory is never to determine a "true" model, which cannot be known and arguably does not even exist. Instead the goal is only to bring clarity and organization to our thinking, and help us to avoid logical fallacies.

1.5 Exam question: Frictionless investment

1. Consider the following model: A firm begins with a capital stock K. It then chooses an investment amount I to arrive at a new capital stock K'. The cost of this investment is $\psi(I, K) \equiv \frac{1}{2}a \times \frac{I^2}{K}$. Then profits will be realized equal to $A \times K'$ where $A > 0$ is a productivity parameter. Then the model ends.

   Derive the regression that is used to test this model under the q theory of investment.

2. Now suppose in addition to the physical capital described in the prior model, the firm also has important intangible capital, It starts the model with amounts K_1 and K_2 of these capital. Then it chooses investment I_1 and I_2, arriving at stocks K_1' and K_2'. It pays separate costs $\psi(I_1, K_1)$ and $\psi(I_2, K_2)$ for each of these investment choices, where ψ is the function defined above. Finally profits are realized according to the Cobb-Douglas function $A \times (K_1')^{\alpha} \times (K_2')^{1-\alpha}$.

   Show that in this case, the regression from the prior question is still a valid implication of the model, but the regression coefficient is different.

3. Now suppose instead that tangible and intangible capital combine additively into a "total" capital stock. As in the prior question, the firm starts with capital stocks K_1 and K_2 and chooses investment I_1 and I_2 to arrive at K_1' and K_2'. The firm pays adjustment costs $\psi = \frac{1}{2}a \times \frac{I_1^2+I_2^2}{K_1+K_2}$, and realizes profits of $A \times (K_1' + K_2')$.

   What type of investment-q regression is predicted by this model?

Solution to exam question 1.5

1. The firm's problem is

$$\max_I \quad A(K+I) - \frac{1}{2}a \times \frac{I^2}{K}$$

The first-order condition is $0 = A - a\frac{I}{K}$, which leads to $\frac{I}{K} = \frac{A}{a}$. We cannot observe A directly, but we do know that $A = V/K'$ where V is the remaining market value AK' after the investment is made. Then we arrive at the testable prediction $\frac{I}{K} = \frac{1}{a} \times \frac{V}{K'}$. So we regress investment rate I/K on Tobin's Q V/K', where V is the firm's enterprise value.

(This is the standard q theory of investment.)

2. Now the firm's problem is

$$\max_{I_1,I_2} \quad A(K_1+I_1)^\alpha(K_2+I_2)^{1-\alpha} - \frac{1}{2}a \times \frac{I_1^2}{K_1} - \frac{1}{2}a \times \frac{I_2^2}{K_2}$$

The first-order conditions are $0 = \alpha\frac{V}{K_1'} - a\frac{I_1}{K_1}$ and $0 = (1-\alpha)\frac{V}{K_2'} - a\frac{I_2}{K_2}$, where $V = A(K_1+I_1)^\alpha(K_2+I_2)^{1-\alpha}$ is again the market value of the firm after the investment decisions are made. Then for physical capital we have the prediction $\frac{I_1}{K_1} = \frac{\alpha}{a} \times \frac{V}{K_1'}$, which is the same regression as before. However, the coefficient will now be $\frac{\alpha}{a}$. Similarly for intangible capital, we can regress I_2/K_2 on V/K_2' and the model predicts a coefficient of $\frac{1-\alpha}{a}$.

(This is the model of Hayashi and Inoue, 1991.)

3. Now the firm's problem is

$$\max_{I_1,I_2} \quad A(K_1+I_1+K_2+I_2) - \frac{1}{2}a \times \frac{I_1^2+I_2^2}{K_1+K_2}$$

The first-order conditions for categories $i = 1, 2$ lead to $\frac{I_i}{K_i} = \frac{A}{a}$. In the current setup, $A = V/(K_1' + K_2')$. So the implied regression for each $i = 1, 2$ is $\frac{I_i}{K_1+K_2} = \frac{1}{a} \times \frac{V}{K_1'+K_2'}$. Compared to the original regression, we now scale enterprise value and investment by "total" capital. If we regress total investment $\frac{I_1+I_2}{K_1+K_2}$ on $\frac{V}{K_1'+K_2'}$ we expect a coefficient of $\frac{2}{a}$.

(This is the model of Peters and Taylor, 2016.)

1.6 Reference: Dynamic investment model

For reference, this section lays out a *dynamic* version of the investment model at the start of this chapter, that also incorporates randomness and uncertainty in productivity at each date, prices of investment goods fluctuating over time, and a few other generalities.[6] If you do empirical work on investment, you will need to study the dynamic setup because it is standard. However, it is not our main focus because we could see the important intuition in the static model, and all other models that we look at in this course will also be static.

- The firm's profits each period are generated according to $\pi(K, A)$, where K is the currently-installed capital stock, and A is a random variable that affects profitability.[7]
- If the firm decides to invest a dollar amount I in new capital, this will impose a cost $\psi(K, I)$ that is convex in I. We will specify ψ more precisely later on, but I want to mention right away that convexity is important.
- There are no taxes. This is just for convenience; we can add in taxes with no qualitative impact.
- At every date t, the firm observes the current capital stock K and profitability shock A, and chooses an investment policy I to maximize the expected present value of future profits $\mathbb{E}[\sum_{\tau=t}^{\infty} \beta^t \pi(K_t, A_t)]$ where β is the relevant discount rate. We don't need to be specific about the dynamics of A (it could be i.i.d, or feature some kind of persistence).
- Existing capital K in each period depreciates at a rate δ before being replenished by any new investment I.

In principle this is a very complicated maximization problem to think about: When the manager chooses I to affect the current capital stock, they do not know what the future profitability A will be, but they do know that their future self will again choose optimally.

However, under standard technical conditions (the usual reference is the textbook by Stokey and Lucas), it turns out that we can characterize the manager's problem by defining a value function V in terms of a *recursive* maximization problem for each time t:

$$V(K_{t-1}, A_{t-1}) \quad = \quad \max_{I_t} \quad \left[\quad \pi(K_{t-1}, A_{t-1}) - \psi(K_{t-1}, I_t) + \beta_t \mathbb{E}_t[V(K_t, A_t)] \quad \right] \tag{1.4}$$

[6] This particular presentation is adapted from Cooper and Ejarque (2003), but other well-known versions are presented in Hayashi (1982) and Erickson and Whited (2000).

[7] This is more general than it seems. We can imagine many other inputs other than "capital" that affect the firm's profits in any given period (e.g. inventories, fleet of vehicles, temporary workers). But if they can be optimized within each period t, then the firm's decision about how much investment to undertake each period will still boil down to a function that looks like $\pi(K_t, A_t)$, where the other inputs have been "optimized out" and absorbed into A_t. This detail is usually mentioned as a footnote in investment theory papers, e.g. footnote 6 in Cooper and Ejarque (2003), so I am following the same tradition here!

subject to the constraint

$$K_t = (1 - \delta)K_{t-1} + I_t \tag{1.5}$$

This type of recursive formulation is called a Bellman equation. The (non-time-varying) function V is called a value function, and its value at any time t is determined by the state variables K_{t-1}, A_{t-1}. If we can find the function $V(\cdot)$ and the choice I_t solving this problem, then we have characterized the solution to the original problem.[8]

As always, we begin the solution with the first-order condition in I_t:

$$\psi_2(K_{t-1}, I_t) = \beta_t \mathbb{E}_t[V_1(K_t, A_t)] \tag{1.6}$$

which has a natural interpretation: At the optimal investment policy, the cost of a marginal dollar of investment (LHS) equals the (discount, expected) benefit of that marginal dollar (RHS).

It is standard at this point to assume a specific functional form for the adjustment cost function. The most common functional form is:

$$\psi(K_{t-1}, I_t) = p_t I_t + \frac{1}{2} a_1 \left(\frac{I_t}{K_{t-1}} - a_0 \right)^2 K \tag{1.7}$$

which implies

$$\psi_2(K_{t-1}, I_t) = p_t + a_1 \left(\frac{I_t}{K_{t-1}} - a_0 \right) \tag{1.8}$$

From here, there are two different basic approaches taken in the literature.

The first approach is to work out something called the firm's "Euler equation" for investment. This involves writing out $V(K_t, A_t)$, the value function measured one date after (1.4), and then working out a specific dynamic connection between these two dates for everything in the problem *other* than V. Unlike the q theory of investment, the Euler equation approach does not rely on measuring V directly. Hence it can work in principle even if the firm is privately held, or if we are not willing to assume that its market valuation is reliable. There is a large literature on this approach including some fascinating econometrics, but it gets quite technical and we will not cover it here.

The other approach is to derive the q theory of investment, following the same logic as we did for the static model in the main chapter. Assume that the profit function π is proportional to the capital stock, $\pi(K_t, A_t) = A_t K_t$. Then one can prove that the value function is likewise proportional, $V(K_t, A_t) = \nu(A_t) \times K_t$ for some ν, which again implies $V_1(K_t, A_t) = V(K_t, A_t)/K_t$. The linearity of the value function is not trivial to prove (the standard reference is Hayashi, 1982, though that model is in continuous time). However, it is often the case that the value function for a dynamic optimization problem will inherit the functional form of the flow payoff functions (in this case $\pi - \psi$).

[8] Cooper and Ejarque (2003) just start from this formulation with their equation (1).

Substitute this into the RHS of (1.6), and (1.8) into (1.6), and rearrange:

$$\frac{I_t}{K_{t-1}} \quad = \quad a_0 \quad + \quad \frac{1}{a_1} \times Q_t \quad - \quad \frac{p_t}{a_1} \tag{1.9}$$

where $Q_t \equiv \beta \mathbb{E}\left[\frac{V(K_t, A_t)}{K_t}\right]$.

This expression for Q_t is complicated by the details of timing and expectations that appear in it. However, these are artifacts of our discrete-time setup, and the specific timing assumptions that we chose. Many models in this literature are presented in continuous time, which yields the more-elegant expression $Q_t = \frac{V(K_t, A_t)}{K_t}$, with everything in principle measured contemporaneously.

The important thing is that we were able to convert the derivative $V_1(K, A)$ to the ratio $V(K, A)/K$ via our assumption that π was proportional to K. This is important because the derivative is impossible to measure, but the ratio is straightforward: It is just the firm's enterprise value divided by its capital stock, and this is straightforward to compute (or at least approximate) in the data.

Chapter 2

Screening models and investment

2.1 Overview

We will first consider problems of asymmetric information. As the name suggests, this is a situation where two parties must interact with each other, but one of them knows something that the other doesn't.

In finance, the interaction is typically the signing of an investment contract, under which an investor provides funds for a risky project that will be operated by a manager. The contract will specify how the project's future (uncertain) cash flows will be divided up, and we assume that the manager knows more than the investor about the probability distribution of those cash flows.

This situation in turn can be separated into two categories, *screening* and *signaling*. Screening is the topic of the current chapter, and refers to a situation in which the uninformed party (in this case the investor) proposes one or more contracts, and the informed party (the manager) can decide which contract to accept, if any. In this situation the investor must anticipate how different types of manager will respond to the choice they are given.

Signaling is the topic of Chapter 3. It describes the opposite case in which the informed party requests a contract, and the uninformed can decide whether to sign it (and possibly can set some of the terms). In this situation the investor must draw inferences about the manager's type based on the choice he already made. These are conceptually very similar situations, but there are some formal differences, as we will see.

Again, the current chapter focuses on screening models. In corporate finance these tend to be models of investment, not capital structure, and they tend to take as given that the contract being used is debt. This is because screening has typically been imagined in the context of a bank lending relationship. By contrast, signaling models tend to focus more on capital structure rather than investment, for reasons we will explain in the next chapter.

We will study screening with a basic investment model: One party (the "entrepreneur") has the opportunity to invest in a risky project, but has no funds, while the other party (the "investor") has the funds but cannot operate the project themselves. In this setting, the natural problem of asymmetric information that arises is that the entrepreneur likely knows more than the investor about the project's characteristics, such as the distribution of its returns.

Of course, every situation in the world features some asymmetry of information. No two people know exactly the same things. We only care about information asymmetries that may likely affect the outcome of the problem. In the setting of investment, information asymmetry would be no problem if the entrepreneur could be trusted to always invest in all positive-NPV projects and ignore all others. So we will look at models in which entrepreneurs not only have more information about the quality of their project, but also may have the incentive to ignore good projects or undertake bad ones as a result.

Selection markets in general

Before describing the specific models in this chapter, it is worth taking a step far back to understand the general idea of a *selection market*. Almost everything we see in this chapter will be a special case of that general intuition.

A selection market is a market with a few key features:

- the consumers of a good impose a cost on the producers,
- that cost, and the utility that the consumer gets from consuming the good, are both heterogeneous *and* correlated across consumers.
- a consumer's type (meaning her cost and utility) are known to her, but unobservable to the producer.

All these features together mean that any given price will attract a *selected* (nonrandom) subset of the overall consumer population, such that the producer's average cost within this subset will not equal the average cost across the entire population of consumers. Producers must take this selection effect into account when setting prices.

The two leading examples of selection markets are insurance markets and credit markets. But in principle the ideas can be applied to any setting where producer and consumer engage in some kind of longer-term relationship beyond an anonymous and instantaneous transaction.

The most interesting thing about selection markets is that they are a prominent setting where the competitive equilibrium outcome will typically *not* be the efficient outcome. To achieve an efficient outcome, we typically need producers to set price equal to their *marginal* cost (that is, the cost imposed on them by the marginal consumer who is indifferent at this price). By contrast, in competitive equilibrium firms must make zero profits, so they set price equal to the *average* cost imposed by consumers who purchase at that price. In a selection market, as observed above, average and marginal costs can be quite different, and it follows that competitive equilibrium typically will not be efficient.

Building on this insight, the literature defines *adverse selection* as a case where willingness to pay is *positively* correlated with cost, and *advantageous selection* as a case where they are *negatively* correlated. Adverse selection leads to an equilibrium quantity that is *below* the efficient amount, while advantageous selection leads to an equilibrium quantity that is *above* the efficient amount.[1]

Einav and Finkelstein (2011) gives very clear intuition for the general economics of selection markets, along with a nice graphical approach for illustrating these concepts. You can understand much of what you need in this area just by studying carefully the figures in that paper. They focus on insurance markets, but you can translate to credit markets almost perfectly by just changing producer $\rightarrow$ lender, price $\rightarrow$ interest rate, cost $\rightarrow$ default, etc. I reproduce their figures below (with some cosmetic changes), but for more detailed discussion it is well worth reading their paper.

We consider the market for some good, defined by a demand curve (consumers' quantity demanded as a function of price), and a marginal cost curve (the cost of producing one additional unit of the good, from any starting quantity). We will assume that only one price can be charged for this good.

Beyond this, we will be deliberately vague about how many producers there are or how exactly they behave. Instead, we will focus on the case of zero profits (price equal to average cost) as our notion of a competitive equillibrium outcome. We will compare this with the efficient outcome (price equals *marginal* cost) to assess the degree of market failure. As mentioned above, the basic point about selection markets is that average and marginal cost will generally be different, so there is no guarantee that equilibrium outcomes are efficient.

Figure 2.1 illustrates the benchmark case in which the marginal cost is constant regardless of the quantity produced. This is the case of *no* selection effect. Since marginal cost is constant, it is equal to average cost. Thus the zero-profit outcome is also the efficient outcome. This seems like a reasonable characterization of most markets, where there is not much difference across consumers in the marginal cost imposed on a producer, and any differences that do exist are not likely to be correlated with the prevailing quantity produced.

Figure 2.2 illustrates the case of adverse selection. This is defined as a marginal cost that *falls* with the quantity produced. That is, the consumers most eager to buy the product are also the most costly to serve. This is a plausible scenario in insurance markets: The person who is most eager to purchase insurance, and willing to pay a high price for it, may be acting that way because they know they have an unusually high expected amount of future claims.

Of course, if possible, you would like to charge different prices to the more and less risky consumers. As we mentioned above, we are assuming in these figures that you cannot do so, because you cannot directly observe the consumer's risk level, although the consumer themselves knows what their risk is. This is

[1] Terminology note: Every economist knows the term "adverse selection," but a surprising number are not familiar with "advantageous selection," even though they are simple mirror images of each other. As a result, you will often hear the term "adverse selection" used interchangeably with "information asymmetry," even though adverse selection is really just one possible case within the broader heading of information asymmetry.

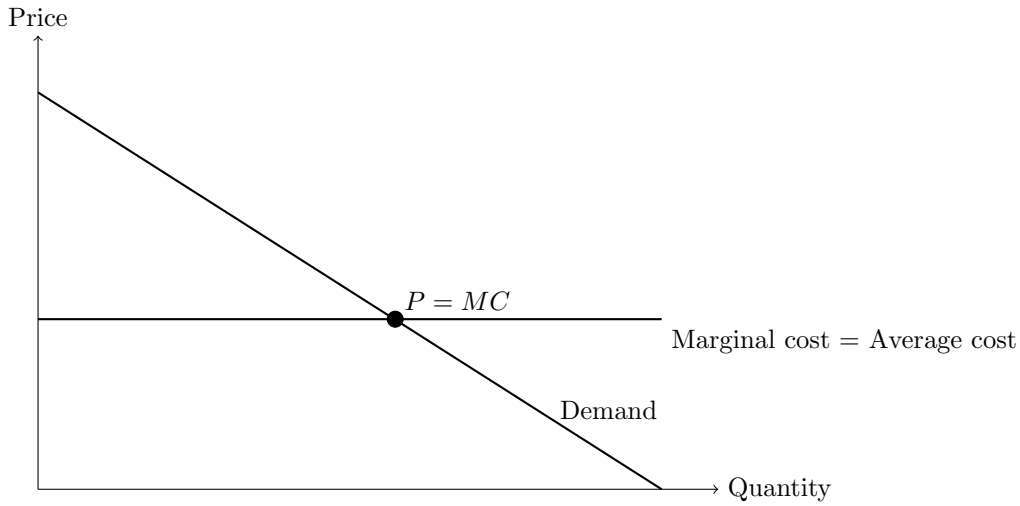


Figure 2.1: No selection effect: Each consumer imposes the same marginal cost. The marginal cost curve is identical to the average cost curve. The efficient outcome $P = MC$ is also the zero-profit outcome $P = AC$.

the sense in which the figures capture a problem of information asymmetry.

As always, the efficient outcome is labeled $P = MC$. At this outcome, the consumers who purchase insurance are exactly the ones who value it more than the cost of providing it to them. However, under adverse selection this will not be the zero-profit "equilibrium" outcome $P = AC$. This is due to the differing behavior of the marginal and average cost curves in the figure.

To be precise, for low quantities produced, marginal cost and average cost are similar (and high). As quantity rises, and marginal cost falls, the average cost falls too but more slowly: it is always an average of the low-cost marginal type who is newly arriving into the market, with the high-cost *inframarginal* types who were already there. Thus the average cost curve is always above the marginal cost curve, with the gap growing as quantity rises.

As a result, the zero-profit outcome $P = AC$ represents too little quantity, at too high a price, compared to the efficient outcome $P = MC$. Adverse selection leads to underproduction and over-pricing of the good being sold. Only the *most* costly consumers will receive the good in equilibrium, while the least costly consumers are left out.

Why can't a producer attempt to serve the unmet demand from the low-cost consumers? If possible, a producer would like to offer them a low-price contract and earn the profits that are being missed by the competitive equilibrium. The problem is that anyone offering the low-price contract cannot get *only* these marginal low-cost consumers. If they undercut the equilibrium price represented by $P = AC$, they will also get all the inframarginal high-cost consumers who were purchasing there, and this will ultimately be unprofitable.

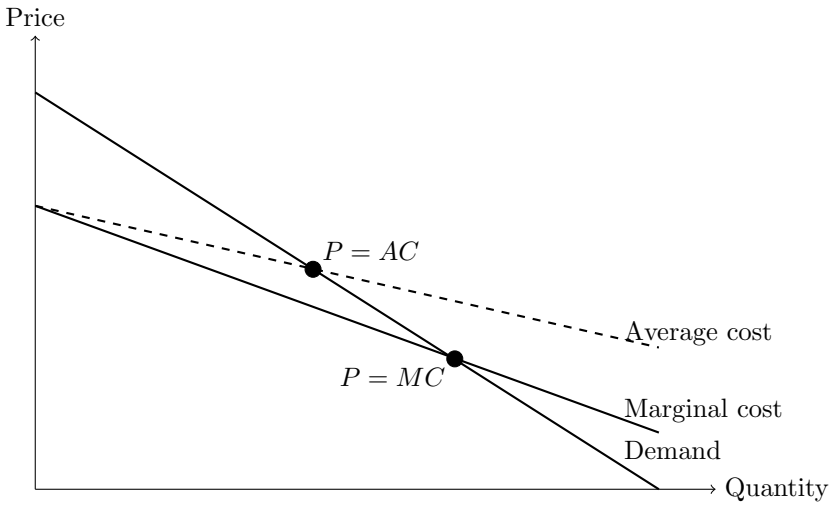


Figure 2.2: Adverse selection: The consumers with the highest willingness to pay are also the most expensive from the producer's perspective. The zero-profit outcome $P = AC$ features too little quantity at too high a price, compared to the efficient outcome $P = MC$.

Again, the basic problem here is the information asymmetry of being unable to distinguish high- and low-cost consumers. Adverse selection adds onto this that the high-cost consumers are most likely to purchase, and then predicts that equilibrium quantity is inefficiently low. In the setting of insurance markets, the belief that adverse selection is widespread leads to concerns that a private market will never be able to provide the efficient quantity of insurance.

Figure 2.3 illusrates the opposite case of advantageous selection. This is the case where the marginal cost imposed by a consumer is *rising* as the size of the market grows. A major theme in modern insurance research is that this scenario is more plausible than was traditionally thought. For example, perhaps the people who most value insurance are generally cautious in every area of their life, and are thus actually lower-risk compared to the general population.

Under advantageous selection, the zero-profit outcome actually exhibits too *much* production compared to the efficient benchmark. Everyone knows that there are some consumers at $P = AC$ who are actually unprofitable for the producer, and should not be consuming, but there is no way to tell who they are. A producer could raise prices to drive out these consumers, and they would earn positive profits, but this would create the opportunity for another producer to undercut them, taking both the profitable inframarginal types as well as the unprofitable marginal types.

Thus we have the mirror image of adverse selection. The key is that with adverse selection, we assumed a *positive* correlation between cost and willingness to pay, while under advantageous selection, we assumed a *negative* correlation.

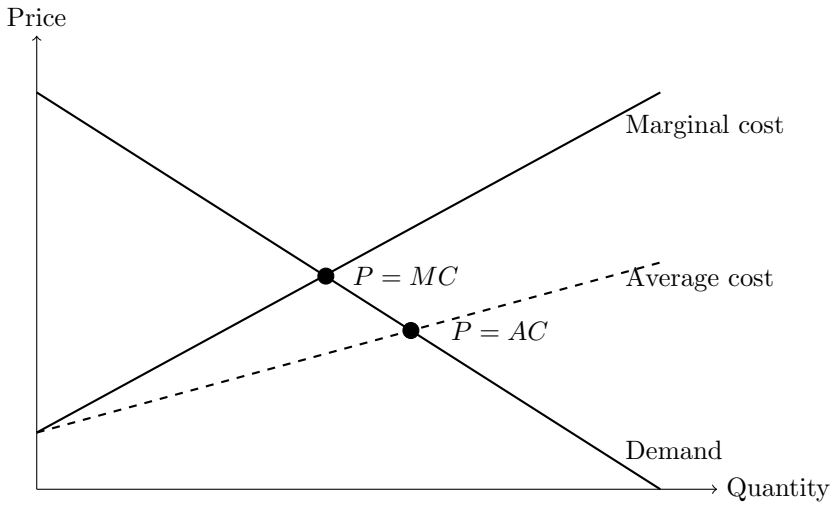


Figure 2.3: Advantageous selection: The consumers with the highest willingness to pay are also the *least* expensive from the producer's perspective. The zero-profit outcome $P = AC$ features too *much* quantity, at too *low* a price, compared to the efficient outcome $P = MC$.

To understand which situation is likely to prevail in practice, we would need to know more about what this correlation looks like across consumers.

In this chapter, we will illustrate all the above intuition in the context of *credit* markets, using a simplified version of the framework in de Meza and Webb (1987). We will modify the framework as needed to capture the main ideas of not only their paper, but also Stiglitz and Weiss (1981) and Bester (1985).

We will also talk a bit about the technical issue of how precisely to define equilibrium. To preview: Everyone tends to agree that competitive equilibrium should feature zero profits for producers, but it is surprisingly difficult to *generate* this outcome from an explicit model of producer behavior within a selection market. Such models often have multiple equilibria, or no equilibrium at all. The literature is not settled on how to think about this issue. For now, the standard approach is just to *impose* zero profits in the definition of equilibrium. We will see this with the papers in this chapter.

2.2 General model of investment

We will lay out a general framework for thinking about how selection effects affect credit markets and investment. Then we will use this framework to illustrate the key results of Stiglitz and Weiss (1981), de Meza and Webb (1987), and Bester (1985). The framework itself is closest to de Meza and Webb (1987).

2.2.1 Environment

Entrepreneurs and their projects

There is a continuum of risk-neutral entrepreneurs with projects. Each project requires investment K. At a later date, project i generates the random cash flow $\tilde{R}_i$, which is equal to R_i^s with probability p_i, and R^f with probability $(1 - p_i)$. Note that R^f is assumed to be constant across borrowers.

Every entrepreneur has identical initial wealth $W < K$ and so must raise financing of $B \equiv K - W$ in order to invest in ther project. Their outside option is to save their wealth at return ρ between the initial and later dates, and thus get a final payoff of $W(1 + \rho)$.

Investors and their financial contracts

Assume that there exist investors who can provide the entrepreneur with capital, but they will only do so using debt contracts, which are defined as follows:

- Initially, the lender provides capital of B to the borrower.
- The debt contract specifies a repayment of $B(1 + r)$ after the project's payoff is realized, where r is an interest rate specified in the contract.
- However, everyone understands that the project payoff $\tilde{R}_i$ is actually a random variable, and there is some probability that it will be insufficient to cover the specified payment. Hence the debt contract really specifies an ex-post cash flow from the borrower to the lender that is also a random variable, $\min\{B(1 + r), \tilde{R}_i\}$.

Assume that lenders cannot observe the type of any borrower who requests a loan. Hence there can only be one interest rate r offered by this market.

Finally, the lender's outside option rate of return is the same as the borrower's. That is, the opportunity cost of making any loan is $B(1 + \rho)$.

Since borrowers and lenders have the same outside option rate of return, this means the *social* opportunity cost of funding any project is $K(1 + \rho)$. Assume that the average project is worth funding from a social perspective, $\mathbb{E}[p_i R_i^s + (1 - p_i)R^f] > K(1 + \rho)$.

2.2.2 Equilibrium definition

Let's first remind ourselves about the basic idea of equilibrium. After setting up a model (who can act, what choices can they make, and what payoffs will they get), we must state an *equilibrium concept*, meaning a set of conditions that (in our opinion) should be satisfied by any legitimate solution to the model. This is a critical piece of the analysis and must be spelled out explicitly.[2]

[2]Some of the older papers we look at in the course are, by modern standards, very loose about describing their equilibrium concept. This is part of what makes them difficult to read.

Strictly speaking, the equilibrium concept is chosen by the authors, and there are often some degrees of freedom in how it is specified. However, the definition will be constrained by convention and norms regarding what is plausible.

The most obvious conditions that we might impose on equilibrium are things like choosing the action that maximizes one's utility. Sometimes we might also impose plausible conditions directly on the *outcome* of the equilibrium, such as that producers will make zero profits, without trying to generate these conditions explicitly from anyone's strategic choices. We will see examples below.

Turning back to the model at hand, we define an equilibrium as:

- an interest rate r offered by the banking sector, and
- a choice by each entrepreneur i of whether to borrow and invest,

that satisfy the following conditions:

- Each entrepreneur i borrows if and only if the expected payoff from doing so exceeds the outside option of saving her wealth, and,
- Lenders expect zero profit from any loan that is taken out.

As mentioned above, this definition mixes an explicit strategic condition in the first point, with a direct restriction on the equilibrium outcome in the second point. The latter is meant to capture the notion of a "competitive" lending market, but again, it is *imposed* as an assumption, not *derived* from any explicit model of lender behavior. This will be important when we get to Bester (1985).

2.3 Pooling with one-dimensional contracts

At this point, the literature has considered two different situations that lead to polar opposite results. Let's look at them in turn.

2.3.1 Stiglitz and Weiss (1981) assumptions and result

Stiglitz and Weiss (1981) assume that all projects have the same expected return, $p_i R_i^s + (1 - p_i)R^f = \bar{R}$. To put this differently, the effect of lowering p_i is a "mean-preserving spread" in the distribution of the random project cash flow $\tilde{R}$. As a result, while there are two quantities R_i^s and p_i that vary across borrowers, there is really only one dimension of borrower heterogeneity, because either of these two quantities perfectly determines the other. An entrepreneur is "riskier" if their project has a lower p_i or equivalently a higher R_i^s.

What type of situation is described by this assumption? The authors have in mind a situation where lenders can judge the expected return of the project just as well as the borrower, and can segment borrowers into different pools by expected return, charging each pool a different interest rate r, but there is some extent to which borrowers are better informed about the *risk* of their projects relative to that average. The model should then be interpreted as describing outcomes within a given pool.

The most important result is:

Proposition 2.1. *If there is sufficient heterogeneity in borrower risk, then equilibrium must feature* less *investment than the first-best level.*[3]

Proof. First recall that we assumed the average project was worth funding from a social perspective, $\mathbb{E}[p_i R_i^s + (1-p_i)R^f] > K(1+\rho)$. Since we have now assumed that all projects have the same average return, this is equivalent to saying *all* projects are worth funding, $\bar{R} > K(1+\rho)$. So it will be sufficient to prove that there is *some* project that does not get funding.

Next observe that the safest entrepreneur (the highest p_i) is the *least* likely to invest. To show this, observe that entrepreneur i borrows if and only if

$$W(1+\rho) \leq p_i[R_i^s - B(1+r)]$$

Rewrite the condition above as

$$W(1+\rho) \leq p_i[R_i^s - R^f] - p_i[B(1+r) - R^f]$$

Write out the condition $\bar{R} = p_i R_i^s + (1-p_i)R^f$, rearrange to $p_i[R_i^s - R_f] = \bar{R} - R^f$, and substitute above to get that entrepreneur i borrows iff

$$p_i \leq \frac{\bar{R} - W(1+\rho) - R_f}{B(1+r) - R^f}$$

In words, the type with the highest value of p_i gets the least payoff from using any given debt contract, hence is the most likely to deviate.

Next write out the zero-profit condition for the lender, which defines r:

$$B(1+\rho) = \mathbb{E}[p_i B(1+r) + (1-p_i)R^f] = R^f + \mathbb{E}[p_i] \times [B(1+r) - R^f]$$

Rearrange this to

$$B(1+r) - R^f = \frac{B(1+\rho) - R^f}{\mathbb{E}[p_i]}$$

Substitute into the denominator above and rearrange to get that entrepreneur i borrows if and only if

$$\frac{p_i}{\mathbb{E}[p_i]} \leq \frac{\bar{R} - W(1+\rho) - R^f}{B(1+\rho) - R^f}$$

For *all* entrepreneurs to borrow would require

$$\frac{\bar{p}}{\mathbb{E}[p_i]} \leq \frac{\bar{R} - W(1+\rho) - R^f}{B(1+\rho) - R^f}$$

where $\bar{p}$ is the greatest value of p_i across borrowers. Hence, there is a sufficient degree of borrower risk dispersion $\bar{p}/\mathbb{E}[p_i]$ such that some projects are not funded (specifically, the safest projects), and the quantity of investment is lower than the first-best quantity. □

[3] This proposition does not precisely correspond to any result in Stiglitz and Weiss (1981). In fact, the presentation in that paper is very different from this chapter, although the essential conclusions are the same. The closest result in the literature to the proposition stated here is actually Proposition 5a in de Meza and Webb (1987), the paper we will study next, where they compare their analysis with Stiglitz and Weiss (1981).

In words, this result tells us that some types will forgo investment if there is sufficient spread between the highest and the average payoff probabilities.

How is this result happening? As you might guess, we have set up a scenario where there is adverse selection in credit markets: The riskiest borrowers are the most eager to borrow. Once we know there is adverse selection, then it is not surprising that there is too little investment in equilibrium, based on the earlier figures.

However, it *should* seem surprising that adverse selection can happen in the first place. Remember that everyone in the model is risk-neutral, including the borrowers, This means that they only care about the *expected* payoff they receive, not the uncertainty of that payoff. Since all borrowers have the same expected payoff, it's not obvious why they would have different levels of eagerness about borrowing.

The way this pattern arises in the model is through the assumption that lenders and borrowers use a debt contract to fund the project. This is a critical and often-overlooked piece of the argument in Stiglitz and Weiss (1981), and it illustrates a general pattern in finance that debt markets can distort incentives in surprising ways. I discuss this point further below (page 36).

There is some useful intuition available by rearranging the final inequality in the proof as

$$\frac{\bar{p}}{\mathbb{E}[p_i]} - 1 \quad \leq \quad \frac{NPV}{X}$$

Here $NPV \equiv \frac{1}{1+\rho}\bar{R} - K$ is the expected social surplus created by each project, in present value terms, and $X \equiv B - \frac{R^f}{1+\rho}$ is the amount of capital that the lender puts at risk. So the problem of "too little investment" disappears as projects become arbitrarily valuable relative to their required external financing ($NPV/X \to \infty$), which is intuitive. A similar cutoff NPV/X also appears in the literature on real options in investment.

2.3.2 de Meza and Webb (1987) assumptions and result

In contrast to the above analysis, de Meza and Webb (1987) assume that R^s is identical across projects, while p_i is still heterogeneous. In one sense this is a simpler assumption than that made by Stiglitz and Weiss (1981), because now there is only one parameter that is different across borrowers. In another sense this setup is actually more complicated, because now there is heterogeneity in both the expected payoff and the risk of that payoff. An entrepreneur with a higher p_i is both "better" and "safer."

One important implication is that different entrepreneurs now have projects with different social values. In particular, some of them may not be worth funding. An efficient outcome should fund the good projects but not the bad. By contrast, in the prior section all projects were worth funding (because they all had the same payoff) so the efficient choice was to fund everything.

The most important result (Proposition 2a in the paper) is:

Proposition 2.2. *Equilibrium must feature* more *investment than the first-best.*

Proof. In this setup, the social value of the project, the borrower's payoff, and the bank's payoff, are all strictly increasing in p_i. Hence, the marginal borrower (the one who is indifferent) must generate negative profits for the bank. Label this borrower by p^*. The indifference condition for this borrower is

$$W(1+\rho) = p^* \times [R^s - B(1+r)]$$

And negative profits for the bank means

$$B(1+\rho) > p^* \times B(1+r) + (1-p^*) \times R^f$$

We can add these together to get

$$K(1+\rho) > p^* R^s + (1-p^*) R^f$$

meaning that the project is inefficient. From this, the set of projects that gets funded in equilibrium includes all efficient projects and some inefficient ones as well. Hence there is more investment than the first-best level. □

How is this result happening? We have now set up a case where there is *advantageous* selection in credit markets: The managers who are most eager to fund their projects are also exactly those with the greatest probability of repaying the loan. Again, this pattern should seem surprising since everyone is risk-neutral here. As in the other paper, it arises as a subtle effect of the fact that (by assumption) projects are funded with debt contracts. Again, see further discussion below.

2.3.3 Comparison and discussion

Stiglitz and Weiss (1981) is much more widely-known than de Meza and Webb (1987), but I hope you can see why the papers should be studied together. As simple as the models are, they have generated a tremendous amount of discussion and further research. In the empirical and policy literature, researchers have discussed how to test for these effects, and what to do in response.

- The result of Stiglitz and Weiss (1981) is often cited by those who feel that credit is too scarce in some setting. If you believe this is happening, it can make sense to *subsidize* lending, for example by a subsidy on interest income, since there is too little lending happening in equilibrium. See Proposition 6 of de Meza and Webb (1987).

- The result of de Meza and Webb (1987) result is cited by those who worry that credit conditions have become too loose and standards are too lax. If you believe this is happening, it can make sense to *discourage* lending, for example by a *tax* on interest income, since there is too *much* lending happening in equilibrium. See Proposition 3 of de Meza and Webb (1987).

  Interestingly, another possible solution in the latter case is to discourage competition and promote market power by the largest lenders! Mahoney

and Weyl (2017) discuss this in detail. I am skeptical that such a policy would be a good idea in practice, but it is interesting for being so contrary to the usual intuition that we should always encourage more competition.

Of course, before making either argument, the most important priority should be to check whether the underlying assumptions of the model indeed seem like an accurate description of reality. In our setting, the most direct thing to examine is the assumption about the underlying distribution of borrower characteristics. In the insurance literature, this is done with "correlation tests" that check whether households that purchase more insurance impose greater or lower costs on insurers *ex post*. This idea was popularized by Chiappori and Salanié (2000) and has been very influential in the insurance literature, most prominently in several important papers by Amy Finkelstein.

In corporate finance, the parallel idea would be to tests whether heterogeneity in project returns is mainly in the second moment (as assumed by Stiglitz and Weiss, 1981) or in the first moment (as assumed by de Meza and Webb, 1987). Unfortunately and surprisingly, there is almost no direct evidence on this question in the setting of borrowing by corporations or small business. One rare exception is Crawford et al. (2018).

Returning to a theory perspective, I will make just a few further comments:

- **Adverse selection via debt contracts**

  Recall the intuition of selection markets (Einav and Finkelstein, 2011):

  - If consumer cost and demand are *positively* correlated (that is, if the consumers with the greatest demand also impose the highest cost on producers), then selection is *adverse* and we should expect the equilibrium quantity of the good to be *less* than the efficient quantity.
  - If consumer cost and demand are *negatively* correlated (that is, if the consumers with the greatest demand also impose the *lowest* cost on producers), then selection is *advantageous* and we should expect the equilibrium quantity to be *more* than the efficient quantity.

  Most research on selection focuses on insurance markets, especially health insurance. In that setting, the type of selection basically depends on consumer characteristics: Are the highest-risk consumers also the most likely to seek out insurance? Many people intuitively say yes (so selection is adverse), but really it's an empirical question. If low-risk consumers are also extremely *risk-averse*, which does seem possible, then selection would actually be advantageous. Empirical research on insurance markets has found many cases where this indeed seems to be true.

  We see a similar tension in the models of *credit* markets that we have just examined: The Stiglitz and Weiss (1981) approach results in too *little* lending compared to the efficient outcome, while de Meza and Webb (1987) results in too *much* lending. As you might guess, this happens because the two models feature adverse and advantageous selection, respectively.

But, unlike with insurance models, here we have assumed that all borrowers have the same, risk-neutral utility function. So there is no mechanical connection between borrower risk and desire to invest. Then, how can adverse and advantageous selection effects arise in this framework?

The answer is that these selection effects arise endogenously due to the use of a debt contract to finance the project. This generates an effect that is ubiquitous in financial economics: The high-risk entrepreneur who borrows to fund her project will only internalize the project's potential positive payoffs, not its losses.[4]

Hence, for any given *average* project payoff $\bar{R}$, the borrowers with higher risk of default will indeed have a stronger desire to invest. In Stiglitz and Weiss (1981) this creates a positive correlation between the borrower's demand and the lender's expected loss rate. Then it is not surprising that we see adverse selection and an inefficiently low equilibrium loan quantity.

In de Meza and Webb (1987), by contrast, the variation across borrowers is in the probability of project success. This variation affects both the mean and variance of project payoff, but roughly speaking, we can say the variation across borrowers is "primarily" in the mean. Then, borrowers with a higher probability of being able to pay off their loans also expect higher payoff from their investments, and we see advantageous selection.

Here is how de Meza and Webb (1987) summarize this intuition:

> "For both the Stiglitz-Weiss model and our model, equilibrium involves entrepreneurs with high-success probabilities subsidizing low-success probability investments. However, there is a crucial difference between the two models in this regard. In our model the marginal project financed has the lowest success probability of those financed, while in the Stiglitz-Weiss model it has the highest. This, of course, explains the asymmetry in the relationship of the respective equilibrium levels of investment to the respective first-best levels."

- **Why debt?** As pointed out above, the use of a debt contract is critical to the results of each model. Then it's natural to ask, could we just solve the problem by using a different contract?

  In the case of SW, this is indeed a serious issue. Many authors have pointed out that equity, not debt, is the "natural" contract to use in the SW setting, and would seem to solve the adverse-selection problem that their model generates. By contrast, in the DW setup it seems that the "natural" contract is debt, so the inefficiency that they predict seems more likely to arise in practice. (I am being loose about the meaning of "natural" because this is beyond the scope of our class, but if you are interested, see the discussion in DW Section III).

[4] To put this differently, her payoff resembles a call option on the underlying project cash flow $\tilde{R}$. This intuition will come up again in Chapter 4.

- **What if the contract includes more terms than just an interest rate?** See the discussion of Bester (1985) in the next section.

- **Equilibrium definition:** Note that we have been loose about stating how many lenders there are, or precisely how they behave. This is deliberate.

  You might have thought that these papers would write down an explicit model of lenders strategically offering loan terms, and reacting to each others' offers, and then would study Nash equilibria of that model. But this approach can become intractable very quickly, and requires us to specify many details that are unimportant to the analysis at hand.

  Instead – as we mentioned earlier – the spirit of the current framework is that we know in advance, to some extent, how we expect a competitive lending sector to behave. Then we can just impose that behavior by making it part of the definition of equilibrium. This somewhat in keeping with the traditional price-taking analysis of competitive markets (e.g. through supply and demand curves), where we assume that each side of the market must simply take prices as given and cannot affect them.

  At the moment, the assumption being imposed seems reasonable enough. But as we will see in the next section with Bester (1985), this approach encounters difficulties as we try to develop the model further.

- **Multiple dimensions of heterogeneity?** In both setups, there is only a single dimension on which borrowers are different. Since they generate polar opposite predictions, it clearly matters a lot which dimension of heterogeneity is the "correct" one.

  As we consider this question, the obvious answer is that in practice, neither model is completely right or wrong. The variation across borrowers in the distribution of their project cash flows must be richer than can be captured by a single parameter. So a natural question is, what would happen in a model that has more than just one dimension of borrower heterogeneity?

  This turns out to be a surprisingly challenging question. The problem is not with mathematical complexity. Rather, the problem is with generating a reasonable definition of equilibrium such that the set of equilibria is neither empty, nor so big as to be meaningless. In this sense the problem is very closely related to the prior point, and I will say more about it following the discussion of Bester (1985) below. But the main focus of the literature, even today, is on models with a single dimension of heterogeneity.

2.4 Screening with two-dimensional contracts

In Stiglitz and Weiss (1981) and de Meza and Webb (1987), loan contracts consisted *only* of an interest rate. Hence there was no possibility of offering different contracts that can "separate" borrowers of different type: If the market offers a menu of different interest rates, everyone will just choose the lowest one

available. Then the only possible outcome was to pool all types together on a single contract, and this is what led to the inefficient outcomes in each model.

In reality, debt contracts always consist of much more than just an interest rate. Many of them feature collateral of one form or another, and corporate loans or bonds will also typically include a long list of covenants restricting the borrower's behavior in various ways, along with penalties for violations. These "non-price" terms vary widely across borrowers and over the business cycle.

One possible interpretation of these patterns is to guess that non-price terms serve precisely to separate borrowers into different contracts according to their type. Indeed, it seems fair to believe that a borrower who is willing to pledge collateral, or accept tighter covenants, is more optimistic about their own repayment prospects than a borrower who is not (all things equal, of course).

Bester (1985) is a model describing how this situation can arise in equilibrium, extending the framework of Stiglitz and Weiss (1981). In his model, loan contracts consist of a promised interest rate, plus an additional cash flow that will be paid to the lender in case of default on that promised payment. He labels the non-price term as "collateral," with the interpretation that the borrower has some extra assets that will be confiscated if the borrower defaults on the promised payment. However, the same specification could be interpreted to nest various types of loan covenants and other loan features.

The important result is that we can sustain a *separating* equilibrium, in which different types of contract are offered, and borrowers of different risk level voluntarily choose different contracts. With this accomplished, each contract can be priced fairly given the borrower who chooses it, and all projects are funded, undoing the problematic result from Stiglitz and Weiss (1981). The lender's behavior of separating borrowers across contracts is known as *screening*.

But why would borrowers voluntarily choose different contracts? Why wouldn't they all choose, for example, the one with a lower interest rate?

In any separating equilibrium, the key is to provide a punishment for choosing the "better" contract, that is more painful for one type than for the other. So, in Bester's model, the contract with a low interest rate will also specify a large amount of collateral (literally, a large transfer to the lender in case of default). This is more painful for the riskier type precisely because they have a higher probability of ending up in the default state, and they know it. So the riskier type chooses to pledge less collateral, and accepts a higher interest rate.

A general feature of both signaling and screening models is that they avoid the worst-case scenario but still are not perfectly efficient. In Bester (1985), we will see that all projects are funded in equilibrium, but, there is still an efficiency loss due to the use of collateral itself. Collateral carries deadweight costs in the model, and serves no fundamental purpose other than to separate borrowers from each other. It would be better from a social perspective if everyone just used their equilibrium contract voluntarily, without the need for pledging collateral. But since this is not sustainable, it may be that the costly "collateral" equilibrium is the best that we can actually achieve.

2.4.1 Bester (1985) assumptions and results

Model

We return to the framework of the previous section, and modify it slightly to reflect the setup of Bester (1985).

For simplicity (not critical to the results), we now imagine there are only two types of entrepreneur instead of a continuum. The entrepreneur types are indexed by $i \in \{a, b\}$. Each entrepreneur can choose a project that requires investment of I. Each has initial wealth W, so in order to invest they need financing of $B = I - W$. Their project return is $\tilde{R}_i$ which is equal to R_i^s with probability p_i, and R^f with probability $1 - p_i$.

We assume that type b is riskier than type a in the same sense as Stiglitz and Weiss (1981): The two projects have the same expected return, but $p_a > p_b$.

Loan contracts specify an interest rate r, and an amount of collateral C. If borrower i obtains the necessary funding B using contract (r, C), then after the project payoff is realized, he pays to the lender a cash flow equal to

$$\min\{B(1+r), \tilde{R}_i + C\}$$

We assume that there is some cost to pledging collateral: At the date the loan is taken, the borrower experiences a loss of kC where $k > 0$. This cost is meant to reflect that firms are restricted in their strategic flexibility as long as their assets are encumbered by collateral claims, and in general, as long as they have committed to restrictive debt contracts.[5] Hence the expected payoff to borrower i from choosing contract (r, C) is

$$\mathbb{E}[\max(\tilde{R}_i - B(1+r), -C)] - kC \tag{2.1}$$

Any contract with $C \geq B(1+r)$ would be risk-free from the lender's perspective and everyone would use it. Therefore we assume that collateral is scarce enough that the only feasible contracts feature $C < B(1+r)$.

We will also assume that collateral is not too costly, specifically that $k < \frac{p_A}{\mathbb{E}[p_i]}$. Clearly, as collateral becomes arbitrarily costly to use, eventually firms will give up on it, and we would revert to the earlier analysis without collateral.

Equilibrium definition

Define an equilibrium as

- a set of contracts that are offered by the lenders, and,
- a choice of contract (or no investment) by each entrepreneur,

[5] More loosely, k can also reflect the costly process of the lender foreclosing on the assets in bankruptcy and taking possession of them. Strictly speaking this is inconsistent with the model as we've written it, because (in order to stay close to Bester's exposition) we assume the cost k must be paid regardless of whether default happens. But you can get very similar results if you instead specify k to be paid only in the event of default.

such that

1. Each entrepreneur's choice maximizes their expected utility.
2. Each contract yields zero expected profit to the lender given those choices.
3. There is no other contract that *could* be offered that would yield a positive expected profit to the lender who offers it, after entrepreneurs are given the chance to switch to that contract.

Item #3 is new. Recall how we imposed a flavor of "competitive" behavior in the previous section by simply defining equilibrium to feature zero profits, without explicitly deriving that behavior. Here we are going one step further: With a richer (two-dimensional) space of potential contracts available, the zero-profit assumption still leaves many possible equilibrium outcomes. Bester tightens those predictions by adding item #3 to the definition of equilibrium. It does feel like a natural property of a "competitive" market, and it rules out many possible equilibria and tightens the predictions of the model. But with this stronger condition, it is not necessarily clear that equilibrium exists at all! See discussion at the end of this section.

Results

Before going through the proofs, it is worth studying Bester's Figure 2, which provides geometric intuition of the argument. The critical thing is that the slopes of the indifference curves are unequal across the two borrowers due to their different payoff probabilities, and the lender's isoprofit slope is different from either one due to the fact that the lender does not internalize the collateral cost k. We will exploit this basic fact to find regions of contracts in $C - r$ space that generate lender profits while attracting only one type of borrower.

Theorem 2.1. *In any equilibrium, the two types use different contracts, and all projects are funded.*

Proof.

- First show that pooling cannot be an equilibrium: Suppose lenders offer a single contract (r^*, C^*) and all borrowers take it. Equilibrium requires that $B(1+\rho) = \tilde{p}B(1+r^*) + (1-\tilde{p})(R^f + C^*)$ where $\tilde{p}$ is the population-weighted average of p_A and p_B. We can show that the slope of the indifference curves for the borrowers are $-\frac{1-p_i+k}{p_i B}$, and since $p_A > p_B$ this slope is steeper for type A. Suppose a lender offers a contract $(\hat{r}, \hat{C})$ that demands more collateral $\hat{C} > C^*$ but a lower interest rate $\hat{r} < r^*$, with

$$\hat{r} \quad \in \quad \left(\quad r^* - \frac{1-p_B+k}{p_B B}(\hat{C} - C^*) \quad , \quad r^* - \frac{1-p_A+k}{p_A B}(\hat{C} - C^*) \quad \right)$$

  and furthermore

$$\hat{r} \quad \in \quad \left(\quad r^* - \frac{1-\mathbb{E}[p_i]}{\mathbb{E}[p_i] B}(\hat{C} - C^*) \quad , \quad r^* - \frac{1-p_A+k}{p_A B}(\hat{C} - C^*) \quad \right)$$

The second range exists thanks to our assumption $1 + k < \frac{p_A}{\mathbb{E}[p_i]}$.

Verify a few things:

Borrower A gets strictly greater utility from $(\hat{r}, \hat{C})$ than (r^*, C^*):

$$\begin{aligned} p_A(R_A^s - B(1+\hat{r})) + (1-p_A)(-\hat{C}) - k\hat{C} &> p_A(R_A^s - B(1+r^*) + (1-p_A)(-C^*) - kC^* \\ -p_A B\hat{r} - (1-p_A+k)\hat{C} &> -p_A Br^* - (1-p_A+k)C^* \\ p_A B\hat{r} + (1-p_A+k)\hat{C} &< p_A Br^* + (1-p_A+k)C^* \\ p_A B(\hat{r} - r^*) &< (1-p_A+k)(C^* - \hat{C}) \\ \hat{r} - r^* &< \frac{1-p_A+k}{p_A B}(C^* - \hat{C}) \end{aligned}$$

Borrower B gets strictly lower utility:

$$\begin{aligned} p_B(R_B^s - B(1+\hat{r})) + (1-p_B)(-\hat{C}) - k\hat{C} &< p_B(R_B^s - B(1+r^*) + (1-p_B)(-C^*) - kC^* \\ -p_B B\hat{r} - (1-p_B+k)\hat{C} &< -p_B Br^* - (1-p_B+k)C^* \\ p_B B\hat{r} + (1-p_B+k)\hat{C} &> p_B Br^* + (1-p_B+k)C^* \\ p_B B(\hat{r} - r^*) &> (1-p_B+k)(C^* - \hat{C}) \\ \hat{r} - r^* &> \frac{1-p_B+k}{p_B B}(C^* - \hat{C}) \end{aligned}$$

So A will deviate to this contract while B will not.

Given this behavior, the lender would get strictly positive profits from offering the new contract. We can deduce this as follows: By assumption the lender earned zero profits at (r^*, C^*). Its isoprofit line, if it continued to serve the entire population, has slope $-\frac{1-\mathbb{E}[p_i]}{\mathbb{E}[p_i]B}$. The new contract sits above that isoprofit line and hence would generate positive profits to the lender if all borrowers continued to choose the contract. In fact we have proved that only A borrowers will use the new contract, and they generate strictly greater profit to the lender than B borrowers do (at any given contract). Hence the lender would make strictly positive profits from offering the new contract, hence the original pooling behavior cannot constitute an equilibrium.

- Similar arguments demonstrate that there cannot be an equilibrium with a single contract used only by one type: If the single contract is used only by type B, then a lender can offer a new contract like the one described above to attract only the A types and make a strict profit. If the single contract is used only by type A, then type B would have the incentive to deviate from this equilibrium and use that contract.

- We conclude that the only possible equilibrium is separating, meaning that both types invest but use different contracts. In a separating equilibrium, both types of borrower obtain all the expected surplus from their

project, since lenders earn zero expected profits conditional on borrower type. Since both projects by assumption have expected surplus that exceeds the borrower's outside option, it follows that both types of borrowers will invest and all projects are funded in equilibrium.

□

Clearly type B will use less collateral and pay a higher interest rate in equilibrium. In fact, type B will pledge no collateral at all. I will not write out the proofs of these facts as they are somewhat tedious, but the intuition is robust and shows up in any model like this one: Collateral is costly and everyone in the model would like to use as little of it as possible. Its only real purpose is for type A to use *more* of it than type B does, as a way of separating. The cheapest way to do this from a social perspective is for A to pledge something and B to pledge nothing.

2.4.2 Discussion

- **The value (and cost) of collateral** The result is impressive: By introducing a non-price loan term, we completely address the concern from Stiglitz and Weiss (1981) that some projects will not get funded. Now all projects are funded, and borrowers voluntarily sort into contracts with different amounts of collateral involved. This has become a standard model to motivate why loan contracts in practice feature more than just an interest rate. However, as Bester acknowledges, collateral itself may carry deadweight costs, so the outcome is not perfect from a social point of view.

- **Importance of lender competition** All the papers in this chapter are intended to be models of highly *competitive* credit markets, with the idea that lenders drive each others' profits down towards zero. This seems increasingly appropriate in the modern era, but of course it has never been exactly true, and for most of history it was very wrong: Loans have traditionally been made in the context of long-term relationships between borrower and lender (e.g. a local bank and a local business), in which the lender has superior information to the rest of the market thanks to their close relationship, that allows them to serve the borrower where outside lenders might not, but also gives them some bargaining power over the borrower. This type of market is inherently *not* competitive and lenders have substantial market power. Market power gives lenders some ability to distort the outcomes derived in this chapter (for better or for worse).

  An illustration is provided by Besanko and Thakor (1987). They consider a setup similar to Bester (1985), and compare equilibrium under two different market structures: A competitive market delivers results with similar intuition to Bester (though with some changes due to other differences in the model setup). A monopolistic market, on the other hand, features no collateral being used at all, and a single contract being offered with an interest rate so high as to deter some borrowers from investing.

Intuitively, the monopolistic lender internalizes all profits and losses in the economy, including the social deadweight costs of collateral usage. He would rather shut down all such costs by allowing no collateral usage at all, and pooling all borrowers onto a single contract, even if this means that many valuable projects go unfunded. This is exactly the behavior that cannot be sustained under perfect competition: An outside lender would always give type A the chance to escape the pooling equilibrium by offering a contract that only she will take (following the process that we detailed in the proof of Theorem 2.1).

Neither perfect competition nor monopoly is an accurate description of credit markets, which are probably always in a state of flux between these extremes. However, the contrasting results that we see for these extreme cases gives us a lot of insight into how competition shapes credit markets.

- **Does equilibrium even exist?** Notice one big thing missing from the results: Any equilibrium that exists must look like the separating equilibrium that Bester describes. But, as he acknowledges in footnote 12, there is no guarantee that an equilibrium exists at all!

  This problem with non-existence of equilibrium is formally identical to the one pointed by Rothschild and Stiglitz (1976) in their earlier (and much more famous) paper on insurance markets:

  - They define competitive equilibrium exactly as in Bester (1985): Insurance providers earn zero expected profit on each contract they offer, and additionally, there is no other contract they could offer that could earn positive expected profit.
  - Under this definition, there cannot be a pooling equilibrium, again for exactly the same reason as in Bester (1985): If there was a single contract used by both types of entrepreneur, then that contract would earn strictly positive profits for the lender on the low-risk types, and strictly negative profits on the high-risk types. A competitor could offer a contract that is marginally preferred by the low-risk type, but not by the high-risk type, and that generates positive profits. They would capture all the low-risk types for themselves, and leave the incumbent firm with only the high-risk types, which generate losses. So the only possible equilibrium is a separating equilibrium.
  - But, as they explain in their Figure 3, it's possible that this equilibrium also fails to exist: If average consumer quality is sufficiently *high*, there will be another contract (labeled γ in the figure) that is preferred by both types to the separating contracts. As reasoned above, pooling on γ also cannot be an equilibrium, so in this case equilibrium would not exist at all.[6]

[6]Note that there is a typo in their figure: the vertical axis should be labeled W_2, the horizontal axis should be labeled W_1.

From these examples you can see that "how exactly to define equilibrium in a screening model" is a general problem and a very difficult one.

How to think about this issue? Everyone agrees that equilibrium should require zero expected profits to be earned on any contracts that are traded in equilibrium. The basic problem arises from saying what should be true about the contracts that are *not* traded in equilibrium. If you place no restriction on these contracts, then a huge range of equilibria can be sustained, many of which are implausible, suggesting that this simple equilibrium concept is "too weak" to make clear predictions. But if you strengthen the equilibrium concept following the idea of Rothschild and Stiglitz (1976) or Bester (1985), you may eliminate *all* equilibria for some parameter calibrations, suggesting that this concept is "too strong."

The literature has struggled to find an appropriate middle ground between these two ideas. A different idea is to build an explicit *strategic* model and study Nash equilibrium and its various refinements, but this approach has also struggled to yield clear conclusions. But some progress is being made. Dubey et al. (2005) and Azevedo and Gottlieb (2017) have both introduced equilibrium concepts that seem to strike a good balance, satisfying general existence theorems while also making fairly sharp predictions.

2.5 Conclusion

It is widely accepted these days that information asymmetries are pervasive in credit markets. Many papers assume that information asymmetry automatically takes the form of adverse selection, which in credit markets leads to an inefficiently low level of investment as described in Stiglitz and Weiss (1981).

However, as we've seen in this chapter, selection effects can distort markets in either direction, and it is not immediately obvious which way is more likely. Unfortunately, even after decades of research and thousands of citations to the papers described in this chapter, the literature is still struggling to come up with clear ways of testing even in which direction information asymmetries distort lending outcomes in practice, let alone how much.

Furthermore, while Bester (1985) gives a convincing rationale for why loan contracts include more than one dimension, there still seems to be a big disconnect between this model and the extreme complexity of commercial loan contracts in practice. (Try reading one sometime!) Intuitively, the Bester model would suggest that there should be as many dimensions in the contract as there are dimension of the information asymmetry between borrower and lender. In reality, there are often dozens of covenants and hundreds of pages of details about how loan contracts work in practice. It seems hard to understand this complexity from the perspective of the models we have at the moment. And the issue seems important, as there is often more empirical variation in these "non-price" terms than in the interest rate that seems much more salient in the typical model of credit markets.

The positive thing about this situation, is that there is much opportunity to contribute to this literature going forward. The main obstacle has been technical issues related to the definition and existence of equilibrium, and the field appears to be making some progress on this at the moment. We may soon see some very important contributions in this area extending the intuition that has remained in place since the classic papers that were covered in this chapter.

2.6 Exam question: Information and investment

Too little investment

Consider the following model based on Stiglitz and Weiss (1981):

- There is a continuum of risk-neutral entrepreneurs, indexed by i.
  Each entrepreneur has a project that requires initial investment of K.
  Each project i, if taken, generates a random cash flow $\tilde{R}_i$, equal to R_i^s with probability p_i, zero with probability $1-p_i$. The discount rate is zero.
- Each project has the same expected cash flow: $p_i R_i^s = \bar{R}\ \forall i$.
  Let $V \equiv \bar{R} - K$ denote the NPV of each project. Assume $V > 0$.
  Assume that $p_i \in [0,1]$, that the distribution of p_i has positive support on all of $[0,1]$, and that average p_i is below a threshold: $\mathbb{E}[p_i] < \frac{B}{B+V}$.
- Each entrepreneur has wealth $W < K$. To fund her project, she invests that wealth, then raises financing of $B \equiv K - W$ and invests that too.
- A bank can provide this financing using a standard debt contract, which specifies an interest rate r, and a repayment of $\min\{B(1+r), \tilde{R}_i\}$.
  Then the entrepreneur's payoff is the residual value, $\max\{\tilde{R}_i - B(1+r), 0\}$.
- The outside option for all agents is a risk-free return of zero.
  That is, the entrepreneur can receive a payoff of W by ignoring her project, and the bank can receive B by refusing to lend to an entrepreneur.

Define a *competitive equilibrium* as

- a single debt contract offered by the bank,
- a choice by each entrepreneur i of whether or not to take this contract,

such that

- each entrepreneur borrows iff this is weakly preferred to her outside option,
- the bank's expected payoff from each loan exactly equals its outside option.

Prove that investment in any competitive equilibrium is inefficiently *low*:

1. Show that in a competitive equilibrium, conditional on r, the entrepreneur's decision whether to borrow can be summarized as a cutoff value of p_i.
2. Based on the above result, identify which entrepreneur always has the *weakest* incentive to borrow and invest, and write out the condition on r for this entrepreneur to borrow and invest.
3. Write out the bank's zero-profit condition, assuming that *all* entrepreneurs borrow, solve it for r, and substitute this into your answer from #2.
4. Show that the resulting condition cannot be satisfied given the model assumptions, and explain why this implies the claim stated above.

Too much investment

Change the model in the previous section to follow de Meza and Webb (1987): Now assume that R_i^s is equal to a fixed value $R^s > K$ across all borrowers.

This in turn means that the expected cash flow $p_i R^s$, and project NPV $V_i = p_i R^s - K$, are *not* fixed values. In particular, $V_i < 0$ for some entrepreneurs.

Prove that investment in any competitive equilibrium is inefficiently *high*:

1. Show that the bank's expected profit on a loan is strictly increasing in p_i. Explain why the following must then be true in competitive equilibrium: Among entrepreneurs i who choose to invest, the entrepreneur with the lowest p_i must generate negative expected profit for the bank.

2. Show that in a competitive equilibrium, the entrepreneur's decision to borrow can also be summarized as a cutoff value of p_i, conditional on r.

3. Explain why the following must then be true in competitive equilibrium: Among entrepreneurs i who choose to invest, the entrepreneur with the lowest p_i must have $V_i < 0$.

4. Explain why all these results imply the claim stated above.
   Give some intuition why the result is different from the prior section.

Solution to exam question 2.6

Too little investment

1. As stated in the model setup, an entrepreneur of type i borrows in competitive equilibrium if and only if her payoff is weakly greater than her outside option, that is,

$$p_i(R_i^s - B(1+r)) \geq W$$

or equivalently

$$\bar{R} - p_i B(1+r) \geq W$$

which we can rearrange to

$$p_i \leq \frac{\bar{R} - W}{B(1+r)}$$

which defines a cutoff in p_i, conditional on r, as desired.

2. Based on the above cutoff, the entrepreneur with the weakest incentive will be the one with the highest value of p_i, i.e. $p_i = 1$. This entrepreneur has $R_i^s = \bar{R}$ and borrows if and only if $B(1+r) \leq \bar{R} - W$.

3. Assuming all entrepreneurs borrow, the bank's zero-profit condition is

$$B = \int_0^1 p_i B(1+r) f(p_i) dp_i$$

where we let $f(p_i)$ denote the density of p_i. The left side is the opportunity cost of making a loan, the right side is the expected cash flow from a loan when all entrepreneurs borrow. This is equivalent to $1 + r = \frac{1}{\mathbb{E}[p_i]}$. Substituting this into the expression from the previous question, the entrepreneur with $p_i = 1$ borrows if and only if $\frac{B}{\mathbb{E}[p_i]} \leq \bar{R} - W$.

4. The final condition above is equivalent to $\mathbb{E}[p_i] \geq \frac{B}{\bar{R}-W} = \frac{B}{B+V}$. We assumed the opposite of this. The implication is that there is no competitive equilibrium in which all entrepreneurs borrow. If there were, the bank in this equilibrium would need to charge an interest rate that earns zero expected profits, but we have just seen that when the bank does this, some entrepreneurs will not in fact borrow. Hence competitive equilibrium (if it exists at all) must feature some projects not being funded.

   We assumed that every project has (the same) positive NPV, so if any projects are not funded, this is an inefficiently low level of investment.

Too much investment

1. The bank's profit from lending to entrepreneur i is $p_i B(1+r) - B$, which is trivially strictly increasing in p_i. Since average profit from all borrowers

must be zero in competitive equilibrium, this implies that the borrower with the lowest value of i must generate strictly negative profits, which are offset by strictly positive profits from other borrowers.

2. Borrowing is profitable if and only if $p_i(R^s - B(1+r)) \geq W$, or equivalently $p_i \geq \frac{W}{R^s - B(1+r)}$.

3. Among entrepreneurs who invest, the one with the lowest p_i must be indifferent, such that $p_i = \frac{W}{R^s - B(1+r)}$, or equivalently $p_i B(1+r) = p_i R^s - W$. We also know from #1 that for this borrower, the bank earns negative profit, $p_i B(1+r) < B$. Combining these together, $p_i R^s - W < B$, that is $p_i R^s < K$, that is $V_i < 0$.

4. Borrowers can be sorted such that efficient borrowers have high p_i and inefficient borrowers have low p_i. We showed that all borrowers above a cutoff value of p_i will borrow in any competitive equilibrium, and those closest to the cutoff will be inefficient. That means the set of borrowers in equilibrium includes all efficient borrowers as well as some inefficient borrowers. From this it is fair to say that equilibrium investment is inefficiently high.

   Intuitively, the Stiglitz and Weiss setup featured *adverse* selection: The entrepreneur with the strongest incentive to borrow, is the worst one (highest risk) from the bank's perspective. In this situation, a competitive equilibrium cannot entice the lowest-risk entrepreneurs into the market, because any bank that lowers interest rates to draw these borrowers in, will also attract all the highest-risk borrowers, who are greater in number and generate negative profits.

   DeMeza and Webb featured *advantageous* selection: The entrepreneur with the strongest incentive to borrow is also the one with the highest repayment probability. Then the above logic reverses. Banks know that there are unprofitable borrowers pooling with the profitable ones, but do not attempt to drive them out by raising rates, because this would allow a competitor to undercut and draw away all the profitable high-p types.

2.7 Screening and investment

Consider the following calibration of the investment model of Bester (1985):

- There are two entrepreneur types, H and L. Each has a project that requires investment of $I = 50$, and will pay off $R = 150$ if successful. Type H is successful for sure, but type L is only successful with probability $p_L = 0.5$. They have no liquid assets to invest, so they need to raise financing of $B = 50$ in order to invest. However, they do have illiquid assets worth $\bar{C} = 40$ that can serve as loan collateral.
- Outside lenders can provide financing for the project using a loan contract. A loan contract (C, r) specifies an amount of collateral $C \in \{0, \bar{C}\}$, and an interest rate r. If a borrower pledges collateral C, they they incur an up-front cost of kC, with $k = 0.2$. After the project payoff is realized, if the borrower cannot make the promised repayment $B(1 + r)$, then the lender seizes the collateral, resulting in a cash flow of $-C$ to the borrower and $+C$ to the lender.
- H and L are equally common. Lenders cannot directly observe types.
- Any capital that is not invested by lenders earns a rate of return $\rho = 0$.

An equilibrium is a set of contracts offered by lenders (at most two), and a choice of contract by each type of borrower, such that each borrower weakly prefers their choice to any other offered contract, lenders earn exactly zero profits on each contract that they offer, and lenders earn weakly negative profits on any other contract that they *could* offer.

1. Show that it cannot be an equilibrium for borrowers to pool on $C = 0$.
   (*Hint*: Solve for the interest rate r that this contract must offer. Show that there is a contract $(\bar{C}, r')$ that lenders could offer with $r' < r$, such that H strictly prefers $(\bar{C}, r')$, L strictly prefers $(0, r)$, and lenders earn positive profits on $(\bar{C}, r')$ when only H switches to it.)

2. Show that it cannot be an equilibrium for borrowers to pool on $C = \bar{C}$.
   (*Hint*: Solve for the interest rate r that this contract must offer. Show that there is a contract $(0, r')$ that lenders could offer with $r' > r$, such that L strictly prefers $(0, r')$, H strictly prefers $(\bar{C}, r)$, and lenders earn positive profits on $(0, r')$ when only L switches to it.)

3. Show that it *is* an equilibrium for the borrowers to separate.
   (*Hint*: In a separating equilibrium, type H must pledge the maximum collateral $C = \bar{C}$ and type L must pledge no collateral $C = 0$. Figure out the interest rates that the lender must charge on each of these contracts in order to earn zero profits, knowing that the borrowers separate in this manner. Then show that at these interest rates, neither type wants to switch to the other type's contract.)

Solution to exam question 2.7

1. If both types pool on $C = 0$, the interest rate r that must be charged for this contract for the lender to earn zero profits satisfies

$$B(1+\rho) \quad = \quad [\pi + (1-\pi)p_L] \times B(1+r)$$

At the given calibration this yields $r = 1/3$.

Now consider the interest rate r' that a lender could offer on the full-collateral contract. We want to show that there is an r' such that:

- H prefers $(\bar{C}, r')$ to $(0, r)$:

$$-k\bar{C} + R - B(1+r') > R - B(1+r)$$

At the given calibration this yields $r' < 0.17\bar{3}$.

- L prefers $(0, r)$ to $(\bar{C}, r')$:

$$-k\bar{C} + p_L \times [R - B(1+r')] + (1-p_L) \times [-\bar{C}] < p_L \times [R - B(1+r)]$$

At the given calibration this is true for any $r' \geq 0$.

- The lender makes strictly positive profits on the new contract when H switches to it: Again, this is true for any $r' \geq 0$.

So if both types pool on $C = 0$, then lenders can earn strictly positive profit on any contract $(\bar{C}, r')$ with $0 < r' < 0.17\bar{3}$. Then pooling on $C = 0$ cannot be an equilibrium outcome.

2. If both types pool on $C = \bar{C}$, the interest rate r that must be charged for this contract for the lender to earn zero profits satisfies

$$B(1+\rho) \quad = \quad [\pi + (1-\pi)p_L] \times B(1+r) \; + \; (1-\pi)(1-p_L) \times \bar{C}$$

At the given calibration this yields $r = 0.0\bar{6}$.

Now consider the interest rate r' that a lender could offer on the zero-collateral contract. We want to show that there is an r' such that:

- L prefers $(0, r')$ to $(\bar{C}, r)$:

$$p_L \times [R - B(1+r')] \quad > \quad -k\bar{C} + p_L \times [R - B(1+r)] + (1-p_L) \times (-\bar{C})$$

At the given calibration this yields $r' < 1.18\bar{6}$.

- H prefers $(\bar{C}, r)$ to $(0, r')$:

$$-k\bar{C} + R - B(1+r) > R - B(1+r')$$

At the given calibration this yields $r' > 0.22\bar{6}$.

- The bank makes a profit on the new contract:

$$p_L \times B(1 + r') > B(1 + \rho)$$

  At the given calibration this yields $r' > 1$.

So if both types pool on $C = \bar{C}$, then lenders can earn strictly positive profit on any contract $(0, r')$ with $1 < r < 1.18\bar{7}$. Then pooling on $C = \bar{C}$ cannot be an equilibrium outcome.

3. If the borrowers separate, then the zero-profit interest rate for type H is $r_H = \rho$, and the zero-profit interest rate for L is $r_L = \frac{1+\rho}{p_L} - 1$. At the given calibration this gives us $r_H = 0$, $r_L = 1$.

   Type H gets payoff of $-k\bar{C} + R - B(1 + r_H)$ by using the full-collateral contract, and would get payoff of $R - B(1 + r_L)$ from the zero-collateral contract. At the given calibration these payoffs are respectively 92 and 50, so type H prefers the high-collateral contract.

   Type L gets payoff of $p_L(R - B(1 + r_L))$ from the zero-collateral contract, and would get payoff of $-k\bar{C} + p_L \times (R - B(1 + r_H)) + (1 - p_L) \times (-\bar{C})$ from the high-collateral contract. At the given calibration these payoffs are respectively 25 and 22, so type L prefers the zero-collateral contract.

   So the pair of contracts $(0, 1)$ and $(\bar{C}, 0)$, along with separation by type L and type H to these two contracts respectively, satisfies all our conditions of equilibrium.

Chapter 3

Signaling models and capital structure

3.1 Overview

In the last chapter we considered screening models, and the menu of contracts that an uninformed investor might offer in order to separate informed borrowers based on their quality. Now we consider signaling models, which are the reverse situation: in which the informed borrower *requests* a contract, and the uninformed investor sets the terms of that offer after drawing inferences about the borrower based on their request.

This chapter focuses on Myers and Majluf (1984), the best-known signaling model in corporate finance, which is a model of capital structure. However, signaling models have been used to shed light on investment policy and many other issues as well. Some examples appear at the end of the chapter.

In a signaling-based model of capital structure, the firm chooses which type of security to issue (debt, equity, or potentially other choices), and participants in that market offer terms, taking into account what they might learn from the fact that the firm chose this contract. Remember that the fundamental question about capital structure, as posed by Modigliani and Miller (1958), is why the decision should matter at all. In the papers at hand, the answer will be that different choices may send different signals that affect firm value.

Intuitively, the argument is that equity is more closely tied than debt to the value of the firm itself. If the firm's insiders thought that the firm was currently overvalued by the market, the best way to take advantage of this would be to issue equity. Investors are aware of this dynamic, interpret equity issuance as a sign that the firm is overvalued, and react by lowering the firm's valuation. In turn, firms anticipate this and avoid equity issuance.

We can generalize this argument to say that firms always use the least "information-sensitive" funding source available. Based on this we should expect firms first to use any available internal cash, then debt, then equity. This

preference ordering is known as the "pecking order" of financing. The idea has become a very influential narrative as to how firms behave.

The most famous paper in this area is Myers and Majluf (1984), which attempts to formalize this argument. They set up a model of issuance by a firm whose insiders know more than the market does, and they conclude that the natural equilibrium outcome that we should expect is a pooling equilibrium, where only debt is issued, never equity. Obviously firms do sometimes issue equity in reality, but their interpretation is that this is a last resort.

However, their original argument left some important logical gaps. A later paper, Noe (1988), fills in these details and spells out exactly the logic behind the argument (and clarifies when it may or may not actually hold). Since this is a theory course, we will study the latter paper in detail, because it also illustrates some important concepts related to equilibrium of signaling games.

The signaling model of Spence (1973)

As with the previous chapter, we first step back to understand the broader context of signaling models in economics.

Spence (1973) is the canonical discussion of signaling. It suggests that people (or firms) may frequently engage in wasteful activities simply to attract better treatment by others. The paper also anticipated some technical concepts in game theory that would only be formalized years later. Unlike many classic papers, it is very well-written, and I highly recommend reading it.

Spence imagines an interaction between two parties, one of whom has superior information to the other. In his original model they are a job seeker, who knows their own "type," and an employer who does not. In a finance context we will reinterpret this to think of a firm raising capital, and the investor who will provide it, with the firm being better informed about its own future prospects.

The job seeker wants a payment from the employer, but the amount that the employer is willing to pay depends on the seeker's type, which the employer cannot observe directly. For simplicity imagine that there are just two types of seeker, labeled H and L. We assume the employer is willing to pay more to type H, whom we think of as the "high-productivity" type.

If this were the full description of the model, we would have a typical problem in the style of Akerlof (1970). Spence now adds another wrinkle: Suppose the worker can take a costly and visible action before the employer makes their offer. In the original model this action is interpreted as obtaining an education.[1]

The interesting thing is that we will not assume education actually has any inherent value. That is, education does make the worker any more productive for the employer. Of course, we would hope that education does increase a

[1] Again, the opportunity to move before the contract is signed is the defining feature of *signaling* models. In the *screening* models of Chapter 2, the separation instead occured through the uninformed party offering a carefully-designed menu of contracts that attract different types to different contracts. Signaling and screening are not mutually exclusive (they could happen in the same model) but the goal is typically to focus on one or the other.

worker's productivity in reality, but Spence shows that workers may pursue education even with any such benefit. Instead, the important thing we assume about education is that it has a *cost* in terms of utility, and crucially that this cost is greater for type L than for type H.

This gives rise to the following possibility: The employer offers a high wage to workers who attain enough education. This required amount of education is set high enough that type H finds the wage worth the trouble of the education, but type L does not. Education then allows H to demonstrate their quality and earn a better wage, but it also represents a deadweight cost to society compared to if employers could simply distinguish H and L directly. This "costly signaling" is a plausible explanation for many real-world situations.

Aside from this clear economic message, Spence (1973) also highlights some technical issues that were partially resolved by later developments in game theory, but are still an ongoing research area, as we will highlight below.

Model

We consider the following simplified version of the model in Spence (1973).[2]

There is a firm hiring a worker. The worker can be one of two types, H and L. From the firm's perspective, the probability that the worker is H is π in the absence of any other information. A worker of type $i \in \{H, L\}$ generates value of V_i to the firm if hired, where $V_H > V_L$.

The worker first chooses an education level $e_i \geq 0$, which may depend on the worker's type i. The firm observes the chosen education level and then offers a wage w. Finally, the worker earns utility of $w - k_i e_i^2$ with $k_H < k_L$.

Equilibrium concept

We define equilibrium as a choice of education level by each type e_i^*, and a wage offered by the firm $w(\cdot)$ as a function of the chosen education level, such that:

- Each type's education choice e_i^* must maximize their utility, taking into account how the firm will respond. That is, $e_i^* \in \text{argmax}_e w(e) - k_i e^2$.
- The firm's wage offer w is set equal to the value it expects to get from the worker. That is, $w = \mu \times V_H + [1 - \mu] \times V_L$, where μ is the firm's posterior probability that the worker is type H given the education level they chose. This is meant to capture competition for workers between firms.
- When the worker's education choice matches the equilibrium choice of one of the worker types, then the firm's posterior probability μ should be calculated according to Bayes' rule, given the prior probability π and the actions chosen by each type in equilibrium.

[2]This version is based on Section 5A of Cho and Kreps (1987, p.208). In this section they label the separating equilibrium as a "screening" equilibrium (p. 210), but by modern conventions this is very nonstandard, and we would instead call it a signaling equilibrium.

The last point is new. An equilibrium now has to specify not only the players' choices e and w, but also the firm's belief μ. We require that μ make sense, to the extent that we can, by enforcing Bayesian reasoning.[3]

In this simple model, there are only a few possibilities of what the firm might observe, and so Bayesian reasoning implies a few specific things:

- If the equilibrium choices of the two types are different, $e_H^* \neq e_L^*$, then we must have $\mu = 1$ whenever a worker chooses e_H^* and $\mu = 0$ whenever a worker chooses e_L^*.
- If the equilibrium choices are the same $e_H^* = e_L^*$, then we must have $\mu = \pi$ whenever a worker chooses this effort level. That is, the firm does not change its mind about the worker.

Both of these points are special cases of the more general reasoning implied by Bayes' rule: The posterior probability that a worker is of a given type i given they chose education level e, should be the probability that a worker of type i chooses e divided by the probability that a worker of *any* type chooses e.

The most important thing to notice about the restrictions on μ is that, in any equilibrium, they apply only to education levels *that are actually taken* by some worker in that equilibrium. In principle a worker could choose some completely different education level, not matching anyone's equilibrium choice, and then the conditions above give us no guidance at all about what μ should be. This will become important below.

Separating equilibrium

The key point in the paper is that there is a separating equilibrium in which type H successfully distinguishes himself: Suppose $e_L^* = 0$, and $e_H^* = \sqrt{\frac{V_H - V_L}{k_L}}$, and $\mu(e) = \mathbb{1}\{e \geq e_H^*\}$. Given the value of μ, type H earns $w = V_H$ and type L earns $w = V_L$. We can easily verify that these values satisfy all our conditions on equilibrium. In this equilibrium, education indeed serves as a valuable signal for type H, allowing her to earn a higher wage than she would otherwise get.

The critical thing in the equilibrium above is that type L sees no benefit to acquiring any education: By earning no education, she gets a payoff of V_L. The value e_H^* is chosen so that L would also get exactly V_L utility from acquiring education level e_H^*, given her high disutility of education k_L. If L acquires any other education level than e_H^*, she will get utility strictly *lower* than V_L. So it is indeed impossible for her to do any better than her equilibrium choice of $e_L^* = 0$. On the other hand, type H strictly prefers her equilibrium utility of $V_H - k_H(e_H^*)^2$ to any other choice she could make.

The fact that H strictly prefers her equilibrium choice may make it sound like she benefits overall from the situation, but this depends on how you look at things. If the two workers could perfectly reveal their types to the employer, then they would get utility of V_L and V_H respectively. Compared to that benchmark, H bears the entire cost of signaling as a way of separating herself from type L.

[3] Today we would describe this as a *sequential* equilibrium (Kreps and Wilson, 1982).

Other equilibria and the intuitive criterion

The equilibrium described above reflects the intuition we expected, but it is not the only one! Suppose we leave the prior case unchanged except to specify $e_H^* = \sqrt{\frac{V_H - V_L}{k_H}}$, which is higher than it was before. We again specify $\mu(e) = \mathbb{1}\{e > e_H^*\}$, now at this new value of e_H^*. This is again a separating equilibrium. However, type H must acquire more education than she did before. She earns strictly less utility than in the earlier case, and now is indifferent between attending college or not.

It may seem odd that H would go along with this. Why can't she just lower her education level to the value we specified before? This was already high enough to deter L from imitating H, so it seems silly for H to go any further than that. However, within the equilibrium as we have specified it here, there is nothing H can do. Decreasing her education level will cause her to be viewed as type L, no matter how unreasonable this may seem, and so she is stuck at the higher value.

There are also pooling equilibria in this model. Suppose each type picks $e_i^* = 0$, and $\mu(e) = \pi \times \mathbb{1}\{e = 0\}$, and $w(0) = \pi V_H + (1 - \pi)V_L$. That is, no one goes to college, anyone who does is perceived as type L, and all workers receive a wage equal to the value of the average worker. Now type H fails to separate herself from L and both receive the same wage, which overvalues L and undervalues H.

In other words, our definition of equilibrium allows for many possible outcomes, including the one we were interested in but also many others, and it cannot distinguish any of them as being more likely than another. How should we interpret this situation?

First, let's highlight why the separating equilibrium we initially described seems special compared to all the others. That equilibrium was the *least-cost separating equilibrium* (LCSE). This is a general term in signaling models for the equilibrium in which the high-productivity type successfully separates themselves from the other type, while paying the lowest possible signaling cost to do so. The LCSE is often viewed as the most natural outcome to expect.

Next, how was it possible to manufacture so many other equilibria? All of them are sustained by a pattern of beliefs (captured in the behavior of μ) that may seem unnatural, but cannot be changed. For example, the pooling equilibrium in which no one obtains education is sustained by a belief that anyone who goes to college is type L. This is an odd thing for the employer to believe, since type L is exactly the one that suffers the most by going to college. But if this happens to be the employer's belief, then it is self-sustaining and leads to an equilibrium, at least according to our definition of equilibrium. In general, the definition of equilibrium gives us flexibility to specify many patterns of beliefs that can sustain many equilibrium outcomes.

Cho and Kreps (1987) propose a way to strengthen the definition of equilibrium, to capture the idea that some belief patterns are unnatural and should not be allowed in equilibrium. Specifically, they propose a rule that they labeled

the *intuitive criterion*:

Suppose some player chooses an action m that is not anyone's equilibrium strategy. (This is the exact situation in which Bayesian reasoning has no impact.) Suppose there is some type of player i for whom the action m could never be preferred to i's equilibrium strategy, no matter how others responded to it. Then the intuitive criterion says that other players should not believe the player who chose m is of type i.

Once you understand what this rule is saying, it should indeed seem quite intuitive. But it's important to note that it is not implied in any way by the other parts of our equilibrium definition. Rather, it is an extra condition that we add into that definition.

To be more precise, the intuitive criterion is usually described as a "refinement" of the set of equilibrium, meaning a rule for picking out some equilibria as more natural than others. There are many different refinements in the literature. Many, like the intuitive criterion, have the spirit of trying to capture natural ideas about agents' reasoning and behavior that are not otherwise imposed by the equilibrium definition.

Returning to the model of Spence (1973), let's see the impact of the intuitive criterion. Think about the payoff to type L from attaining education levels above $\sqrt{\frac{V_H - V_L}{k_L}}$, which was the value of e_H^* from the LCSE that we described earlier. This payoff depends on the wage that the employer offers to a worker with such a high education level. However, even in the best possible scenario where the employer assesses this worker as type H and offers a wage of V_H, type L still ends up with utility lower than V_L from having attained such a high education level. Since type L always gets utility of at least V_L in each equilibrium we have described, we conclude that in any equilibrium, attaining such a high level of education could never be a desirable action for L, regardless of how the employer responds. Then the intuitive criterion tells us that, in any equilibrium, a worker who attains an education above the given threshold *must* be viewed by the employer as type H. But once this is true, none of the equilibria other than the LCSE can survive: Type H can choose the LCSE education level and end up better off than her equilibrium payoff. The intuitive criterion thus provides a powerful argument for why the LCSE should indeed be the expected outcome in a signaling model.

3.2 The "pecking order" in capital structure

With this background in the general economics of signaling models, we now turn to discuss how signaling ideas have been applied in research on capital structure.

3.2.1 Myers and Majluf (1984)

Some stylized facts are as follows:

- Firms finance investment most often with retained cash; then with debt issuance; and least often with equity.

- The announcement of an equity issuance by a public firm results, on average, in a decline in the stock price. Announcements of stock repurchases have the opposite effect, increasing the stock price.[4]

- Equity issuers and purchasers both frequently argue that equity issuance has the negative effect of "diluting" the firm's insiders. This is often argued as a reason for the falling stock price pattern mentioned above. There can be no such "dilution" if equity is issued at a price that everyone regards as fair (another example of the argument of Modigliani and Miller, 1958). Instead, the argument implies that, somehow, outside investors value the firm's newly-issued equity less than the firm's insiders do.

One possible explanation for all of these facts is that the firm's insiders have more information than outsiders about the firm's future profitability, and that outsiders know this and pay attention to the firm's actions as a way of inferring something about that private information.

In fact, in the presence of information asymmetries, the decision to sell *any* security has to be, on average, a negative signal about the quality of the firm, because insiders will always have a stronger incentive to sell that security when their information is that the firm is worth less than the market currently thinks. But we would naturally expect that this effect is stronger for equity, which seems like it should be more strongly "sensitive" to the firm's inside information. This is the basic argument that Myers and Majluf (1984) try to capture.

Taken to the extreme, one could imagine that these issues affect not only the firm's choice of how to finance an investment, but also which investments get made in the first place. Maybe some bad investments are taken, or some good investments are ignored, as the firm plays the game of trying to avoid negative signals and send positive ones. This "real effect" was a major focus of the screening-based papers in the last chapter, and it does appear to some extent in the signaling literature too. But the bigger focus in this literature is on questions of capital structure and security design, that is, what contracts will be chosen to finance a *given* investment.

Model

- The firm has assets in place worth A. The firm has the opportunity to invest I in a project that will generate a cash flow of $I + B$ for a net payoff of B. The firm has slack on hand S, so it needs to raise financing of $E \equiv I - S$ in order to invest.

- For the moment, assume the firm can only raise financing by issuing equity.

[4] It's less clear whether any such patterns hold for debt issuance, because that happens so much more frequently: it's difficult to find any announcements of debt issuance that were plausibly a "surprise" and not already anticipated by the market beforehand.

- Assume that the firm's actions will be chosen to maximize the value V^{old} realized by old shareholders, who remain passive throughout the model and do not take part in any new issuance by the firm.[5]

- If the firm does nothing, $V^{old} = A + S$. If it issues equity and invests, $V^{old} = \frac{P'}{P'+E}(A + \underbrace{S + E}_{I} + B)$. Here P' is the market value of the firm's old shares if issuance happens.

Information asymmetry can prevent investment

Just looking at the payoffs in the model, we see that old shareholders want the firm to issue and invest if and only if

$$S + A \leq \frac{P'}{P' + E}(E + S + A + B)$$

which is not guaranteed. The firm invests if the new project value B is above a linear function of A. (Figure 1 in Myers and Majluf (1984) plots that function.)

This result already demonstrates how information asymmetry can prevent a valuable investment from happening. But to be clear (and as the paper acknowledges in Section 2.5), it largely reflects the obvious intuition of Akerlof (1970), as applied to the setting of investment (although with the slight twist that the investment decision interacts with the value of existing assets-in-place; see page 202). The interesting part is really the next section: what if the firm can issue debt instead of equity?

Argument for pooling on debt

Rewrite V^{old} in the case of equity issuance as $V^{old} = S + A + B - \Delta E$. Here $\Delta E \equiv E_1 - E$, where $E = I - S$, and E_1 is the newly issued shares' market value at $t = 1$. And similarly write V^{old} in the case of debt issuance as $V^{old} = S + A + B - \Delta D$.

Then Myers and Majluf (1984) present the following argument for why debt will be the only security issued in equilibrium:

[5]This assumption deserves some discussion. We typically imagine that the firm acts to maximize shareholder value (for the most part). But here there are two different categories of shareholder: those who already own equity at the start of the model, and those who might buy into any new issuance. Furthermore, their interests may not always align: If new investors overpay for the newly-issued equity, that is obviously bad for them but benefits the old investors. Then, it is not clear what it means to "maximize shareholder value."

The assumption being made here is that the firm acts on behalf of *existing* investors only, meaning it will take advantage of new investors for the sake of old investors if it can, and indeed this is the key dynamic behind the whole model. Most researchers accept this as plausible, and in fact many papers assume something similar, but some find it controversial. If the firm blatantly takes advantage of new investors, it might face liability for securities fraud. Even short of that, it's not clear if the manager would want to act this way, or if old investors can force them to. These concerns are duly acknowledged in Myers and Majluf (1984, p.189).

- The incremental payoff to old shareholders from issuing equity is $B - \Delta E$, and from debt is $B - \Delta D$. Hence, equity issuance requires $\Delta E < \Delta D$.

- But, "based on option pricing theory" we should expect that $|\Delta D| < |\Delta E|$.

- Combining these, equity issuance can only happen if the firm believes $\Delta E < 0$, in which case new investors should not buy at all.

MM summarize their argument as follows:

> "Thus, our model may explain why many firms seem to prefer internal financing to financing by security issues and, when they do issue, why they seem to prefer bonds to stock. This has been interpreted as managerial capitalism – an attempt by managers to avoid the discipline of capital markets and to cut the ties that bind managers' to stockholders' interests. In our model, this behavior *is* in the stockholders' interest."

But, while intuitive, this argument has a serious gap...

The problem with the argument

Myers and Majluf (1984) simply *assert* that $|\Delta D| < |\Delta E|$ based on option pricing models. But these are not models of information asymmetry. They tell us that equity value reacts more to news about the firm's value. They do *not* tell us that equity value reacts more when investors update their beliefs in response to the firm's actions.

These may sound like the same thing, but the latter situation is much more nuanced: An investor's reaction to a manager's given choice depends on how they think the manager *would* choose in any possible scenario, which we call the manager's strategy. In turn, the manager's strategy depends on how they think the investor would react to any given action. These definitions are circular, and depend on possibly self-fulfilling beliefs, rather than just the fundamentals of the situation. This makes it less obvious what will actually happen.

For example, what prevents the following situation? "All firms issue equity; any firm who issues debt is believed to be bad; as a result, no firms issue debt." This situation contradicts the argument of Myers and Majluf (1984), but there is nothing obviously wrong with it: all firms are acting optimally given what the market believes, and the market's beliefs are consistent with firms' behavior.

Next we will discuss Noe (1988), which investigates these issues in more detail. We will see that there is indeed an argument to be made that the equilibrium described above is "unreasonable." But it takes a good bit more detail to see exactly what that argument is. Along the way, we will see the ingredients of a rigorous model of signaling and beliefs, illustrating some of the general ideas we saw earlier with Spence (1973) and Cho and Kreps (1987).

3.2.2 Noe (1988)

We will focus only on Sections 1 and 2 of Noe (1988). There are two players in this game: the firm, which requests a type of security; and "the market," which responds by offering terms or by rejecting the request.

Model

- The firm's assets in place will generate a cash flow $t_1 > 0$ at time 1.
  An investment of I at time 0 will yield $t_2 \geq 0$ at time 1.
  Refer to $t \equiv (t_1, t_2) \in \mathbb{R}^2$ as the firm's type, and $q(t) \equiv t_1 + t_2$ as its quality.
  Type is drawn from a finite number of values $t \in T = \{t^1, t^2, ..., t^2\}$.
  Assume that no two types t have the same quality.

- The firm knows t. The market does not know t, but knows that it is drawn from a prior distribution $p(t)$.

- At time zero, the firm announces a message $m(t) \in \{d, e, c\}$: respectively, request debt financing, request equity financing, or reject the project.

- Then the market updates its beliefs from the prior distribution $p(\cdot)$ to a posterior distribution $\mu(\cdot|m)$, and chooses a response $r \in R(m)$ that defines the terms of financing: complete rejection n, a face value k if the firm requested debt, or a fraction α if the firm requested equity.

- The firm acts to maximize the expected value of the time 1 cash flows to its existing insiders, which is denoted $u(t, m, r)$, and is equal to $(1-\alpha)q(t)$ if equity financing is used, $\max(0, q(t) - k)$ if debt, and t_1 if the project is not funded (because either the firm or the market rejected it).

Equilibrium concept

Given this setup, Noe (1988) defines an equilibrium to consist of messages $m^*(t)$, responses $r^*(m)$, and investors' posterior beliefs $\mu(\cdot|m)$, such that

1. firms choose optimally, that is, $m^*(t)$ maximizes $u(t, m, r^*(m))$ for all t,

2. the market receives zero expected profit from projects that are financed, where its expectation is evaluated according to the beliefs μ,

3. the market accepts any project that it believes to have positive-NPV, where its expectation is evaluated according to μ,

4. the beliefs specified by μ are consistent with the market updating by Bayes rule: for each m that is actually selected by some type t according to m^*, $\mu(t^j|m) = 0$ if m is not the equilibrium action specified by m^* for type t^j, and $\mu(t^j|m)$ is equal to $p(t^j)$, divided by the sum of $p(\cdot)$ across all types t that are actually specified by m^* to choose m.

Let's carefully understand each point of this definition:

#1 is obvious: the firm acts to maximize its expected cash flow.

#2 and #3 impose "competitive" behavior on the market.[6] Importantly, they spell out the role of investor beliefs: These beliefs are summarized by a distribution μ over potential borrower types, which can vary depending on the message m that the borrower has sent. Contract terms (α or k) then adjust so that the contracts being used in equilibrium offer zero expected profits, where the expectation is evaluated against the distribution μ.

#4 imposes the natural Bayesian restriction on the market's beliefs μ in order for them to be consistent with "rational" reasoning. To be precise, for any action that is actually taken in equilibrium by some type of firm, the market's belief about a firm taking that action has to be consistent with Bayes' rule, given firms' equilibrium strategies.[7]

But note again that Bayes' rule cannot say anything about beliefs in response to an action that is *not* taken in equilibrium, and hence #4 is silent about such actions. This will become important soon.

Initial results

Noe (1988) starts with a result that seems to echo Myers and Majluf (1984):

Lemma 3.1 (cf Lemma 1 in Noe, 1988)**.** *If some type issues debt in equilibrium, then no type strictly prefers equity, and the face value of debt is I.*

Proof. If a type succesfully issues debt (as opposed to just requesting it), this means that the market does not reject requests for debt financing. The face value that the market offers must be at least I. Then any firm with a negative-NPV project knows that if it requests debt, it will be approved, and its payoff will be worse than if it had rejected its project completely. So in equilibrium such firms cannot optimally choose debt, and the only debt issuers are those whose projects will generate returns greater than I. Since the market knows this, equilibrium requires that the face value of debt is exactly I.

Next, suppose (by contradiction) that some type strictly prefers equity to its other choices, given the terms offered by the market. Since each type can always issue debt that only requires paying I for sure, the type that prefers equity must feel that it is paying less than I in expectation to the market. Since the market must earn exactly I on any equity issued in equilibrium, there must be *another* type that also issues equity and expects to pay *more* than I to the market. But then this type should strictly prefer to issue debt instead at face value I, a contradiction.

[6] Note that we do not explicitly model how that competition works, an approach that should be familiar from the discussion in Chapter 2. Giammarino and Lewis (1987) show that some conclusions change if equity is issued through a process of negotiation, instead of into competitive financial markets, which echoes the point made by Besanko and Thakor (1987) about the importance of perfect competition in Bester (1985).

[7] Again, this is the approach taken in the sequential equilibrium concept of Kreps and Wilson (1982) as well as in the earlier analysis of Spence (1973). The use of Bayesian updating in games of incomplete information originated with Harsanyi.

To put this in simpler terms, when two types of different quality both issue equity, and debt is available, one of them must be making a suboptimal choice. Hence the only possible equity issuance is by the lowest-quality type, at a value that makes them indifferent to either debt issuance or rejecting their project completely (depending on whether that project is negative-NPV). □

But now we can begin to see what was missing from Myers and Majluf (1984): There can *also* be an equilibrium in which *no* types issue debt, and in which the market believes that anyone who does issue debt must be low-quality, which reinforces the equilibrium behavior.

To illustrate, Noe provides an example on page 337 in which there are two firms who both issue equity. One firm's project is positive-NPV and the other's is negative-NPV. Clearly the "good" firm is being undervalued by the market, due to the pooling behavior of the "bad" firm. Nonetheless the good firm cooperates with the equilibrium and issues equity. If it issued debt, the market would assume it was a bad type and reject its request for financing.

One can also set up examples where the good firm does not invest at all, if investment value is small enough relative to the market's undervaluation of its equity, which resembles the first result of Myers and Majluf (1984) – even though debt is available, which was supposed to fix that problem.

How can such opposite results come out of the same model? Again, this is a general difficulty with equilibria in signaling models: The players' beliefs must be specified as part of any equilibrium, and they play a critical role in determining which equilibria are sustainable. But there is a surprising amount of flexibility in what those beliefs must be, regarding actions that are *not* taken in equilibrium. This makes it difficult to rule out a wide range of equilibrium outcomes, even some that appear to contradict one another.

When we investigate these situations carefully, we often find that some of the equilibrium outcomes require beliefs that seem "unreasonable." In the example given above, it seems unreasonable for the market to believe that a debt issuer is worse than an equity issuer, because, in any possible equilibrium the bad type is the one that would suffer *most* from successfuly issuing debt. The intuition here is indeed exactly as suggested by Myers and Majluf (1984).

But as clear as this seems, there is nothing in the definition of equilibrium so far that can capture this reasoning. Our only restriction on beliefs is that they follow Bayes' rule (item #4 in the definition). But if you think carefully about this restriction, you will see that it has no content for actions that are not actually chosen in equilibrium by any player. In other words, the problem arises specifically in pooling equilibria, when we consider what beliefs are permissible regarding a deviation to an action that is *off the equilibrium path*.

In our setting, when all firms choose to issue equity, the issuance of debt is an off-equilibrium action. Then, the definition of equilibrium allows the market to believe anything whatsoever about a firm that deviates and issues debt. So in particular, it is permissible for the market to think the worst possible thing about such an issuer. And if that is the market's belief, then firms will rationally not issue debt. There is never any opportunity to correct the market's belief,

as unreasonable as it may be, and so the equilibrium is sustained. "Indeed, by utilizing the 'right' off-equilibrium beliefs, virtually any pattern of security issuance can be the outcome of a Nash equilibrium" (Nachman and Noe, 1994).

One could argue we should take such situations seriously as a possible description of empirical reality. However, the usual response in the literature is instead to strengthen our equilibrium definition, in such a way as to capture our intuition for why the "unreasonable" equilibrium should not happen.

Refining equilibrium with the Cho-Kreps intuitive criterion

The general idea is that if someone deviates from a pooling equilibrium, other players should consider which type benefits the most or the least from such a deviation, and the answer should affect their belief about the deviator.

In our setting, in an equilibrium where all firms issue equity, this intuition should rule out the belief that a firm who deviates and issues debt is low-quality, because that is exactly the type of firm that would suffer from this deviation.

The most common implementation of this idea is the **intuitive criterion** of Cho and Kreps (1987). This is a "refinement" of equilibrium that requires beliefs after a deviation to assign zero weight to any type for whom that deviation is *strictly* dominated. Noe (1988) adopts the Cho-Kreps intuitive criterion by adding the following condition to the definition of equilibrium:

- (MCK) For any given equilibrium, and any *off-equlibrium* message m', beliefs in response to that message should assign zero weight to any types who are strictly worse off if (1) they choose m' and (2) investors actually accept this offer on terms consistent with their beliefs.

If you think about this, you can see why it feels like a reasonable thing to require, but is not implied by the earlier definition of equilibrium.[8,9]

Then Noe proves the following result, which takes us one step closer to the original idea of Myers and Majluf (1984).

Lemma 3.2 (cf Lemma 2 in Noe, 1988)**.** *Suppose that at least two types have access to positive-NPV investment opportunities. In any equilibrium satisfying (MCK), some type* ***requests*** *debt financing.*

Proof. Suppose (by contradiction) that there is an equilibrium where no type requests debt financing. (MCK) forces the market to believe that any firm who deviates and *does* request debt, must have a positive-NPV project. Therefore, the market will accept such requests at face value I, and any negative-NPV type will not request debt. Now consider the positive-NPV firms. Both must

[8] "MCK" in Noe's paper stands for "modified Cho-Kreps." The subtle difference from the Cho-Kreps definition is the addition of requirement (2). This is necessary in Noe's setting because investors could simply ignore the deviator's message, in which case the deviation is irrelevant to payoffs and the refinement has no bite.

[9] It is standard to call the intuitive criterion a *refinement* of the set of equilibria, meaning a rule for selecting some equilibria as more natural than others. However, you could also call it a *condition* to be added to the *definition* of equilibrium. The difference is just semantic.

be issuing equity, as either one would prefer debt issuance to rejecting their project, given the face value I. But for the market to break even, the firm with the greatest quality must be undervalued given the market's terms, meaning that it must be paying more than I in expectation. This firm should strictly prefer debt to equity, a contradiction. □

This the most important step for you to understand in Noe's analysis. In virtually every nontrivial signaling model, there is some process of "refining" equilibria by restricting off-equilibrium beliefs, as in this example. The Cho-Kreps intuitive criterion is the most common approach to this.[10]

However, one tedious detail remains before we are completely done formalizing the pecking-order intuition.

Refining equilibrium by ruling out weakly dominated strategies

Even with the result of the prior Lemma, it is still possible to find equilibria violating the intuition of Myers and Majluf (1984). Suppose a bad firm type requests debt *and is refused*, while a good type requests equity and is accommodated. Once again, the market will believe that a firm requesting debt is a bad type, sustaining the equilibrium. (Noe provides an example in footnote 9.)

Why isn't this situation ruled out? The request for debt is no longer an *off-equilibrium* action, so (MCK) no longer has any role in ruling out this behavior. Instead, the market *correctly* believes that a firm requesting debt is a bad type. Its belief is consistent with firm's equilibrium strategies, as required by condition 4 of the original equilibrium definition.

Are there any grounds on which we can argue that *this* is an implausible equilibrium? The weird thing now is that the bad type's request for debt is only justified because she strongly believes that this request will be *ignored*, so that she thinks her choice really makes no difference at all. If there is any chance at all that the request might actually be granted, even by accident, then she would be better off not requesting debt after all.

To say this more formally, her equilibrium strategy is *weakly dominated*: A strategy of not requesting financing yields a payoff that can never be lower, regardless of the market's response, and will be strictly higher for some response.

So Noe (1988) introduces one final condition for equilibrium: that no type plays a weakly dominated strategy (according to a definition he provides at the top of p.341). Then he is finally able to demonstrate the original pecking-order intuition with the following two results: The first shows that there is no

[10] Another common approach is based on the concept of "divinity" as introduced by Banks and Sobel (1987). The most common example of this is known as "D1." The basic idea is to strengthen Cho-Kreps such that beliefs put all weight on the types that were *most likely* to have deviated, rather than simply ruling out those for whom deviation was strictly dominated. Outside the literature specifically on signaling games, there are also many other equilibrium refinements, and a large literature attempting to microfound them and reconcile them with each other in some kind of unified framework.

equilibrium in which any firms issue equity, and the second shows that there *is* an equilibrium in which all firms issue debt.[11]

Proposition 3.1 (cf Prop 1 of Noe, 1988)**.** *Suppose that at least two types have valuable projects. In any equilibrium satisfying (MCK), and in which no type plays a weakly dominated strategy, no type strictly prefers equity financing.*

Proof. By contradiction: Suppose there exists an equilibrium where at least two types have positive-NPV projects, in which some type strictly prefers requesting equity to requesting debt. From Lemmas 3.1 and 3.2 it must then be the case that some type requests debt, and that requests for debt are rejected by the market. Rejection by the market happens if and only if $\mathbb{E}_\mu[q(t)|m] \leq I$. From the equilibrium restriction on beliefs, this means there is at least one type t' who requests debt and actually is weakly-negative-NPV, $q(t') \leq I$. But for that type, requesting debt would be weakly dominated by skipping the project entirely: We imagine if the market actually accepted requests for debt on terms that would seem fair given its beliefs. Given the market's beliefs as written above, it would have to set $k \geq q(t)$. This means the borrower's payoff to choosing debt is $\max(0, q(t') - k) = 0$ whenever lenders accept the debt proposal in a sequentially rational way, whereas the borrower could simply reject the project and earn $t_2' > 0$. □

Lemma 3.3 (cf Lemma 3 in Noe, 1988)**.** *Suppose that at least one type has a valuable project. Then there exists an equilibrium satisfying (MCK), in which no one plays a weakly dominated strategy, and in which good types issue debt and bad types reject their projects.*

Proof. We simply construct such an equilibrium:

- $m^* = d$ if $t_2 > I$, otherwise c
- $r^*(d) = k^* = I$
- $r^*(e) = I/(\min_t q(t))$ provided this ratio is less than 1; otherwise $r^*(e) = n$ (reject).
- $\mu(t^j|d) = p(t^j)/\sum_{t:t_2>I} p(t)$ for all $t^j : t_2 > I$; $\mu(t^j|d) = 0$ for all other t.
- $\mu(t^j|e) = \mathbb{1}\{t = \arg\min_t q(t)\}$

In words: Firms with positive-NPV projects request debt, other firms reject their projects. Firms who request debt are priced according to the average among all firms with good projects. Firms who request equity are valued as though they were the lowest-quality type and priced accordingly. We can easily verify that this satisfies all the conditions of equilibrium. □

[11] In general it is *not* standard to rule out equilibria just because they feature weakly dominated strategies. There are settings where it feels reasonable, like this one, but also settings where it can lead to eliminating equilibria too aggressively. The literature has put a lot of effort into formalizing arguments for when this does or doesn't "make sense," and one should really try to appeal to such an argument in practice. Here, we can forgive this step of the analysis, because Noe is really trying to reverse-engineer the original intuition of Myers and Majluf (1984), and demonstrate the kind of thinking that is necessary to obtain their result.

Thus, we have fully clarified the assumptions (or at least, a sufficient set of assumptions) behind the pecking-order argument for debt pooling. You can see why it was important to develop the apparatus of game theory and signaling models that we have introduced in order to demonstrate this result rigorously.

The rest of Noe (1988) focuses on extending this result to models where even the firm's insiders aren't perfectly informed about the firm's cash flows, and this line of thought is continued in the security-design framework of Nachman and Noe (1994). These topics are important (and the second paper is very well-known), but we will not have time to cover them in our course.

3.3 Conclusions

As we've seen in this chapter, models focused on investment have typically thought about screening, while models focused on capital structure have mainly thought about signaling. Much of the same intuition comes up in either context.

A general challenge with both literatures, and the entire topic of information asymmetry, is that it seems possible to sustain almost any behavior in equilibrium, if the nature of the information asymmetry problem is just right, if agents agree to focus the right signal and follow the right off-equilibrium beliefs, etc. Since it's inherently difficult to test any of the underlying assumptions, this literature is always at risk of being an exercise in which we can rationalize almost anything we see as being the outcome of some information asymmetry.

Hence the most important contribution to be made at the moment, is to further nail down exactly which outcomes seem plausible and which one don't. We have seen some flavor of this with the intuitive criterion of Cho and Kreps (1987) in this section, and with the discussion of correlation tests and equilibrium definitions towards the end of Chapter 2, but much remains to be done.

For finance researchers, a particularly fruitful path forward is to pay attention to developments *outside* of finance: New tools are continually being developed by economists working on health, IO, and other topics, and there are always opportunities to import their ideas into the finance setting.

3.4 Exam question: The pecking order

This question investigates the "pecking order" logic of Myers and Majluf (1984), using a simplified version of the model from Noe (1988):

- There are two players, a firm and the market.
  Everyone is risk-neutral and there is no discounting.
- The firm has a type $s \in \{H, L\}$, which it knows but the market does not.
  The market believes that the firm's type is H with probability π.
- The firm already has assets-in-place worth C_s.
  It also has the opportunity to pursue a new investment.
  The new investment costs I, and will produce Y_s at a future date.
- The firm can ignore its investment, or request funding from the market.
  The market responds by setting the terms of the investment contract.
  (The details about this are given in the following sections.)
- If the firm ignores its investment, then its payoff is equal to C_s.
- If the firm obtains funding for the investment, then that funding pays for the cost I, and the firm's value grows to $C_s + Y_s$, which is then divided between the market and the firm according to the terms of the contract.

Assume the following:

1. $C_H > C_L$ (Assets in place are more valuable for type H.)
2. $Y_H > I > C_L + Y_L$ (Investment is valuable for type H, but not for L.
   In fact, the entire value of firm L after investment is less than I.)
3. $\frac{C_H}{C_L} > \frac{Y_H}{Y_L}$ (There is more information asymmetry about assets-in-place than about the new investment opportunity.)
4. $\frac{C_H+Y_H}{\mathbb{E}[C_s+Y_s]} > \frac{Y_H}{Y_L}$, where $\mathbb{E}[C_s + Y_s] \equiv \pi(C_H + Y_H) + (1-\pi)(C_L + Y_L)$.
   (Information asymmetry is severe, that is, π is relatively small.)

An example calibration is $C_H = 10$, $C_L = 1$, $Y_H = 8$, $Y_L = 4$, $I = 6$, $\pi = 0.1$.

3.4.1 Only equity

In this section, the only available investment contract is equity. If the firm issues equity, then the market pays for I, and demands in return a fraction α of the future value $C_s + Y_s$. The payoff to the firm is the residual, $(1-\alpha)(C_s + Y_s)$.

An *equilibrium* specifies

1. a choice by each firm type H and L whether to issue equity or do nothing;
2. the market's posterior probability $\tilde{\pi}$ that an equity issuer is type H;
3. the market's response α to an equity issuance.

These are required to satisfy the following:

(a) Each firm chooses optimally, given the market's response.
(b) Market beliefs about *equilibrium* actions align with the choices in (1). That is, if only H issues equity, $\tilde{\pi} = 1$; if only L, $\tilde{\pi} = 0$; if both, $\tilde{\pi} = \pi$. (If neither issues equity, then this condition places no restriction on $\tilde{\pi}$.)
(c) The market earns zero expected profits from any equity investment:

$$I = \alpha \times [\tilde{\pi}(C_H + Y_H) + (1 - \tilde{\pi})(C_L + Y_L)]$$

Show that in any equilibrium, type H will not invest, using the following steps:

4.1 First show that there cannot be a separating equilibrium in which type H issues equity and invests, while type L does nothing.

4.2 Next show that there cannot be a pooling equilibrium in which both types issue equity and invest.

(*Hint*: For each of the above cases, first determine what is the zero-profit value of α that the lender would set if everyone behaved as described. Then just show that, given this α, some firm would choose to deviate from the equilibrium.)

3.4.2 Introducing debt

Now firms can also request debt. Assume that the face value of the debt is fixed at I, and the market responds to the request by simply accepting or refusing it. If the market accepts, it pays for the investment I. Then the firm repays the face value I if it is able, and otherwise hands over its entire value $C_s + Y_s$.

Now an equilibrium specifies

1. A choice by each firm type to issue equity, request debt, or do nothing;
2a. The market's posterior probability $\tilde{\pi}_E$ that an equity issuer is type H,
2b. Another posterior probability $\tilde{\pi}_D$ that a firm requesting debt is H,
3a. The market's response α to an equity issuance,
3b. The market's response (accept or refuse) to a request for debt.

These are required to satisfy the following conditions:

(A) Each firm chooses optimally, given the market's response.
(B) Market beliefs about equilibrium actions align with the choices in (1), in the same sense as condition (b) from the prior section.
(C) The market earns zero expected profits on any equity given $\tilde{\pi}_E$.
(D) The market accepts a request for debt if and only if $\tilde{\pi}_D = 1$.

4.1 Explain the intuition behind the "pecking order" argument:
Why might we expect that type H will now invest using a debt contract?
(*Hint*: Would type L ever *want* to obtain debt financing?)

4.2 Consistent with the intuition, there is now an equilibrium where type H invests using debt, while type L does nothing. Write out the condition for L to prefer its choice in this equilibrium to equity issuance, and solve it for α. Combine this with the lender's zero-profit condition on equity issuance, and rearrange to get an upper bound on $\tilde{\pi}_E$. Show that this bound is strictly less than 1. Why does it make sense that $\tilde{\pi}_E < 1$?

4.3 *Contrary* to the intuition, there is also an equilibrium where no firm invests. This equilibrium must feature $\tilde{\pi}_D < 1$. Why does $\tilde{\pi}_D < 1$ seem intuitively unreasonable? Why doesn't (B) prevent this situation?

Finally, we "refine" the set of equilibria with the following requirement:

- If a type s would be better off doing nothing than *obtaining* debt, then,
  (1) type s cannot *request* debt in equilibrium, and
  (2) $\tilde{\pi}_D$ must assign zero probability that a firm requesting debt is type s.

4.4 Explain how this restriction rules out the equilibrium from question **4.3**.

Solution to exam question 3.4

4.1 As suggested in the hint, we note that in such an equilibrium, α would be equal to $\frac{I}{C_H+Y_H}$. Given this, assumptions 2 and 3 plus some algebra show that type L would choose to invest, contradicting the equilibrium.

A better approach is to show from Assumption 3 that L always has a strictly stronger incentive than H to take any given equity contract α. This immediately rules out the equilibrium described here, and is useful intuition for the whole problem.

4.2 In such an equilibrium, α would be equal to $\frac{I}{\mathbb{E}[C+Y]}$.
Then type H would *not* invest, contradicting the equilibrium, because

$$\frac{C_H+Y_H}{\mathbb{E}[C+Y]} > \frac{Y_H}{Y_L} > \frac{Y_H}{I} \implies C_H > \frac{\mathbb{E}[C+Y]-I}{\mathbb{E}[C+Y]} \times (C_H+Y_H)$$

4.1 The problem with H issuing equity was that L would try to do the same.
The market then had to penalize the valuation of equity,
to the point that H no longer found issuance worthwhile.
Here, debt is unattractive to L: It would cost the firm its entire value.
Hence, *intuitively*, we should not expect L to attempt debt issuance,
and the market should be able to assume a borrower is type H.

4.2 L would switch to equity issuance if $\alpha < \frac{Y_L}{C_L+Y_L}$. Hence, in equilibrium we must have $\alpha \geq \frac{Y_L}{C_L+Y_L}$. Solve the zero-profit condition for α and substitute into the left side of this inequality, then solve the result for $\tilde{\pi}_E$, and

$$\tilde{\pi}_E \leq \frac{\frac{I}{Y_L}-1}{\frac{C_H+Y_H}{C_L+Y_L}-1}$$

which is strictly less than 1 by $\frac{I}{Y_L} < \frac{Y_H}{Y_L} < \frac{C_H+Y_H}{C_L+Y_L}$. This makes sense because it is the low type that should want to use equity, not the high type, now that debt is available. The terms of equity financing depends on market perceptions about the firm's value, which opens the possibility that type H is undervalued by the market, while the terms of debt financing are fixed, so there is no chance that H is undervalued.

4.3 This seems unreasonable because it implies a positive probability that a type L firm issues debt. As discussed above, this would be a mistake from the perspective of type L. However, condition (B) does not rule this out, because the request for debt financing is not an equilibrium action.

4.4 It forces us to set $\tilde{\pi}_D = 1$. Then the market expects to break even on debt financing and will approve a request for debt. Then type H is willing to invest using debt, and we are left with the equilibrium of question **4.2**.

3.5 Exam question: Signaling through retention

This question derives the main results of Leland and Pyle (1977). Their idea is that an entrepreneur might try to signal the quality of her project through the amount of equity that she retains in that project.

A risk-averse entrepreneur with initial wealth W_0 has a project that requires funding of $K < W_0$. She could pursue the project on her own, or could sell a stake in it to a risk-neutral investor. We assume that the investor uses an equity contract. So the entrepreneur chooses a fraction α to retain, and sells the rest $1 - \alpha$ to the investor.

The project will generate a cash flow $\mu + x$. Initially, the entrepeneur's and investor's information about them is as follows:

- No one observes x, but they know that has mean zero and variance σ_x^2.
- The entrepreneur observes μ. The investor only knows that $\mathbb{E}[\mu] < \infty$.

After the initial decisions, both μ and x are revealed, and payoffs are realized.

From this, the investor's ex-post payoff will be $(1-\alpha)(\mu + x)$. Assume that financial markets are competitive, so the risk-neutral investor pays $(1-\alpha)V(\alpha)$ for her stake, where $V(\alpha)$ measures her subjective valuation of the firm after observing the entrepreneur's choice of α, that is, $V(\alpha) \equiv \mathbb{E}[\mu|\alpha]$.

Then, the entrepreneur's terminal wealth, if she invests, will include the funds she raises by selling equity, plus her payoff from the share that she retains:

$$W_1 = W_0 - K + (1-\alpha)V(\alpha) + \alpha(\mu + x)$$

Assume that she maximizes a mean-variance utility function of terminal wealth:

$$U(W_1) = \mathbb{E}[W_1] - \frac{1}{2}\, b\, \mathrm{Var}(W_1)$$

We will focus on a "separating equilibrium" in which the entrepreneur's choice of α completely reveals μ. That is, in such an equilibrium $V(\alpha^*) = \mu$, where α^* is the *equilibrium* value of α.

In fact, there are many separating equilibria. So we will specifically focus on the "least-cost separating equilibrium," in which the investor believes that any entrepreneur who chooses $\alpha = 0$ has a worthless project, $V(0) = K$.

1. Use the following steps to show that in the least-cost separating equilibrium,

$$\mu \;=\; K \;+\; b\,\sigma_x^2 \left[\ln\left(\frac{1}{1-\alpha^*}\right) \;-\; \alpha^* \right]$$

   where α^* is the entrepreneur's equilibrium choice of α.

   (a) Write out a first-order condition for α that implicitly defines α^* and $V(\cdot)$ in terms of each other.

   (b) Impose the key assumption of a separating equilibrium as described above, then integrate with respect to α, to arrive a function $V(\alpha)$.

(c) Finally, impose the key assumption of a *least-cost* separating equilibrium, set $\alpha = \alpha^*$, and again impose $V(\alpha^*) = \mu$, to arrive at the equation above.

(d) Prove that this equation defines a unique optimal α^* for each μ.

2. Give some economic intuition for the tradeoff that entrepreneurs face in choosing α. Why do they benefit from a higher α? Given this benefit, why does it make sense that in equilibrium each entrepreneur chooses the value of α^* corresponding to their value of μ as defined by the above equation, and does not attempt to "mimic" someone with a higher μ by choosing a higher α?

3. Show that the set of funded projects is efficient. That is, the entrepreneur undertakes her project if and only if a social planner would choose to fund it.

4. Still, in an important sense the equilibrium outcome is *not* efficient. Why?

Solution to exam question 3.5

1. (a) Write $U(W) = (1-\alpha)V(\alpha) + \alpha\mu - \frac{1}{2}\,b\,\alpha^2\sigma_x^2$, and then

$$0 = -V(\alpha^*) + (1-\alpha^*)V'(\alpha^*) + \mu - b\,\alpha^*\sigma_x^2$$

   (b) Impose $V(\alpha^*) = \mu$ and the above becomes $(1-\alpha^*)V'(\alpha^*) = b\alpha^*\sigma_x^2$.

   (c) Divide through by $(1-\alpha^*)$ and integrate both sides $d\alpha$ to get

$$V(\alpha) = C + b\sigma_x^2\left[\ln\left(\frac{1}{1-\alpha}\right) - \alpha\right]$$

   up to a constant of integration C. Set $C = K$, evaluate at $\alpha = \alpha^*$, and replace $V(\alpha^*) = \mu$.

   (d) The RHS is a strictly increasing in α^* on $\alpha^* \in (0,1)$, hence invertible, which implies a unique α^* satisfying the equation for any value of μ.

2. The equilibrium beliefs function $V(\alpha)$ is strictly increasing in α as we just observed. Hence an entrepreneur who chooses a higher α will convince the investor that she has a better project. This has a positive net effect starting from an equilibrium where investors correctly value the project, if we just look directly at the first two terms in $U(W)$ above: Starting from an equilibrium where $V = \mu$, an increase in α and hence V will unambiguously increase the first two terms. So the effect is unambiguously positive through these terms.

   But, the third term explains why the entrepreneur hesitates to set a higher α. This requires him to bear a greater amount of risk. The investor rationally revises her beliefs in a way that takes into account how painful this is to the entrepreneur, and in turn the entrepreneur lives with her revealing choice of α rather than try to mimic anyone else.

3. We want to show that $\mu - K > \frac{1}{2}b(\alpha^*)^2\sigma_x^2$ (the entrepreneur invests) if and only if $\mu > K$ (investment is efficient). In equilibrium $\mu - K = b\sigma_x^2\left[\ln\left(\frac{1}{1-\alpha^*}\right) - \alpha^*\right]$, so we want to show on $\alpha^* \in [0,1]$ that $\ln\left(\frac{1}{1-\alpha^*}\right) - \alpha^*$ is strictly positive iff if it is strictly greater than $\frac{1}{2}(\alpha^*)^2$. Note that both expressions equal zero at $\alpha^* = 0$. So it is sufficient to observe that, for every $x \neq 0$, the derivative w.r.t α is strictly greater for the first expression than the second (that is, $\frac{1}{1-x} - 1 > x$).

4. If the entrepreneurs could reveal μ to the investor, then the equilibrium outcome would be Pareto superior to what we have derived, because every entrepreneur with a valuable project could just choose $\alpha = 0$ and sell the entire project to the investor for price μ. The same set of projects would get funded, but entrepreneurs would have higher utility by bearing less risk.

However, this equilibrium is not sustainable with asymmetric information. The investor cannot simply rely on the entrepreneur to report her valuation μ, because even those with low valuations would report high values of μ. Instead, the investor forces the entrepreneur to "prove it" by choosing higher α. This is purely a costly signal in that it does not benefit the investor and only hurts the entrepreneur. But it is used anyway, because it hurts a bad entrepreneur more than a good one.

Chapter 4

Hidden action models

4.1 Overview

Hidden actions in general

In this chapter we will study models that incorporate **hidden actions**. This refers to situations where two parties sign a contract, and then one of them takes an action that will influence both. For our purposes, we again imagine an investor funding a project and a manager who operates that project. The manager must make a decision after the project is funded that affects the overall payoff of the project, and hence the cash flows to each party.[1]

In many settings, the manager will have incentives at the later date to deviate from maximizing the total value of the project. For example, perhaps they will just steal all the money if they are able. Then the interesting question is what kind of agreements the two parties can reach ahead of time to influence the borrower's incentives later on, encouraging them to act in alignment with maximizing total value.

At a high level, this situation shares many features with the screening and signaling models of prior chapters. In each case there is potentially a misalignment of both information and incentives between the two parties. In screening and signaling models, the greater emphasis is on the asymmetry of information, because this asymmetry exists at the date the contract is signed. In hidden-action models, the greater emphasis is on incentives, because the important decision is made after the contract is signed, and the two parties have symmetric information at all dates where any choices are being made.

[1] Terminology: It is very common in finance and accounting to use **agency problems** or **moral hazard** interchangeably with hidden action. Strictly speaking the first term is slightly more general and the second is slightly narrower, but when you see any of these terms used, you should typically assume they all mean the same thing.

Hidden actions and corporate governance

The incentives studied in hidden-action models are a first-order issue for understanding corporations' financing and investment behavior. Although the shareholders of the corporation nominally control it through voting rights and possibly other means, from a practical perspective they have very little influence over what the corporation actually does. Then how can they trust the corporation with their money? What measures might the corporation take to reassure them, and, what are the potential costs of taking those measures (or failing to take them)? Can these issues explain the patterns that we observe in practice regarding the firm's decisions, contracts, etc.? Might they still leave us with an inefficient outcome, or even push us farther away from efficiency?

Although these questions are old, and were famously explored in books like Berle and Means (1932), the finance and economics literature of the mid-20th century mostly disregarded them as a detail to be figured out by practitioners.

That changed in the 1970s, and especially with Jensen and Meckling (1976). This is one of the most-cited finance papers of all time, and initiated decades of research on corporate governance. It considered several different angles from which incentive issues might affect the firm's choice of investment, capital structure, managerial compensation, etc. The formal modeling ideas that the paper introduced were mostly not new. Instead, its major contribution was to draw them all together, and argue that incentive issues are in fact a central concern, not a side issue, for understanding corporate finance.

To understand why this was such an influential idea, it helps to also understand some context of the US corporate sector in the 1970s. This was a time of very large, diversified conglomerates that seemed intuitively to face little pressure from the outside world. Apologists for this situation, especially the corporate executives themselves, would argue that such organizations can achieve efficiencies and economies of scale that would not otherwise be possible. But many observers were becoming skeptical of this viewpoint around the time that Jensen and Meckling (1976) was published. The paper provided a justification for these concerns, by arguing why businesses may become less efficient as they become larger and more removed from their shareholders.

The 1980s turned out to be a time of pushing back hard in the opposite direction: Many large conglomerates were separated into smaller, more-focused businesses, and this was often done forcibly over the objections of management thanks to the rise of the buyout industry. Jensen and Meckling (1976) and other papers in this literature provided a theoretical justification for this activity.

Turning back to theory, we can list a few different specific types of incentive problem that have been considered in the literature:

- the manager's decision about *how much effort* to exert;
- the manager's decision about *which projects* to invest in;
- the manager's decision whether to *accurately disclose* information to investors, and to *follow through* on promises made to them in the past.

The first two were considered in Jensen and Meckling (1976) and we will discuss them in detail in this chapter. The third will come up in the next chapter, along with the topic of "security design."

We will start by considering some of the ideas in Jensen and Meckling (1976). Their overall narrative is to point out a problem that arises with equity financing based on incentive concerns; then point out a separate problem that arises with *debt* financing, also based on incentives; and conclude by sketching a model in which the firm sets its capital structure by balancing these two problems against each other. However, like some of the other classic papers we have studied, this one falls short of being a rigorous theory by modern standards. So after overviewing some of its main ideas, we will depart from it and develop those ideas using frameworks from other papers.

4.2 Agency problem of equity: Effort decision

Section 2 of Jensen and Meckling (1976, p.312) highlights the "agency costs of outside equity" in a problem of *effort choice.* They consider the manager's decision about how much effort to exert at work. They point out that if the manager's only incentive to exert effort is some equity ownership of the firm, then, the manager's effort choice will be below first-best anytime they hold less than 100% of the firm's equity, and this issue will become worse as their share of the firm's equity gets smaller.

We can demonstrate the idea formally with the following simple analysis based on first-order conditions. This analysis actually does not explicitly appear in Jensen and Meckling (1976), but it is standard in the literature and if you read their Section 2 carefully you can see that this is basically their argument.

- Suppose the manager chooses an effort level e that can directly increase the firm's profits $\Pi(e)$, but imposes a convex personal disutility cost $c(e)$ on the manager. Clearly the optimal effort level from a social perspective sets $\Pi'(e) = c'(e)$. How can the manager be rewarded and incentivized to choose this optimal effort level?

- At small firms, the manager essentially owns the firm and directly internalizes Π. As the firm grows, the manager will sell off equity stakes to other investors, but will still retain some for himself, and so is still affected by Π at the margin. The conventional wisdom is that this equity-based incentive will help make sure that the manager acts in the interest of the firm's profits. But how effective is this incentive scheme?

- Suppose the manager has retained a fraction α of the firm's equity, and this ownership stake provides his only incentive to exert effort. Then the manager's problem is to choose $\max_e \alpha\Pi(e) - c(e)$, and he will respond by setting $\alpha\Pi'(e) = c'(e)$.

- If we assume that profits are increasing and weakly concave in effort, as seems natural, then we can conclude that the manager will set effort

strictly below the first-best level. The intuition is simple: The manager always internalizes the full marginal cost of his effort, and only the fraction α of the marginal effect of his effort on firm profits.

- The effect can be quite large, and it gets stronger as α shrinks. Once we are talking about a mature corporation with a highly-dispersed investor base, in which the company's decision makers hold only a small fraction of equity, we could plausibly believe that managerial effort will be orders of magnitude less than the first-best amount.

This analysis is indeed very simple, but the overall point is an important one: Equity-based compensation can align the manager's incentives in the same *direction* as increasing firm value, but these incentives are far too weak in practice to guarantee the right *level* of effort. We should not be surprised if managers of large corporations slack off and take it easy, potentially destroying large amounts of value for their investors. This was indeed a common cynical perspective on management behavior at the time the paper was written. There has also been a large research literature on the potential for managers to stop exerting effort when they feel that their job is secure (e.g. Bertrand and Mullainathan, 2003).

The next important insight is that equity investors should fully understand what's going on when they decide whether to invest. So, it is actually firms and their insiders who suffer from the issue highlighted here, as equity investors will be reluctant to provide them capital, or will offer unattractive valuations, in anticipation of these issues.

4.3 Agency problem(s) of debt: Project choice

Given the results of the prior section, plus the well-known tax advantages of debt, Jensen and Meckling (1976) ask the following:

> [W]hy don't we observe large corporations individually owned with a tiny fraction of the capital supplied by the entrepreneur in return for 100% of the equity and the rest simply borrowed? At first this question seems silly to many people [...] We have found that oftentimes this question is misinterpreted to be one regarding why firms obtain capital. The issue is not why they obtain capital, but why they obtain it through the particular forms we have observed for such long periods of time.

In other words, why do firms issue equity at all? Why don't they just use debt for all external financing needs, and let the management own close to 100% of the firm's equity, in order to have strongly-aligned incentives?

Of course, if you asked the management of a growing company why they don't fund that growth through debt, they would answer that it would be prohibitively expensive, or simply impossible, to raise that much debt financing through bank loans or bond issuance. So the real question is why the economics of financial markets have led to that outcome. Why should equity be cheaper?

The answer must be some problem with debt issuance, that offsets the problem with equity issuance pointed out in the prior section, and that at some point shifts the firm's preference to be for equity instead of debt.

Following this intuition, Jensen and Meckling (1976) in their Section 4 undertake to highlight such a problem with debt issuance. Moreover, they undertake specifically to highlight an *incentives*-based problem with debt issuance, so that their argument is unified with their earlier analysis of equity contracts.

The specific problem that they highlight with debt financing is known as **asset substitution**. The intuition is that when the firm has a large amount of debt in place, equity investors will develop a preference for risky projects.[2] The manager, acting in shareholders' interest, might then choose to invest in very risky actions even if they have negative expected value. This ultimately destroys value for debt investors, and the distortion of incentives becomes more severe as the firm has a higher debt load. In anticipation of this issue, lenders may refuse to provide more credit to the firm at some point, or may charge exorbitant rates for any credit that they do provide. This could explain why firms perceive that they cannot grow through large amounts of debt financing, and must issue equity despite its shortcomings.

Asset substitution has a mirror-image problem known as **debt overhang**: Due to their preference for risk, shareholders may also ignore projects with *positive* expected value if they aren't risky enough. This problem is not explicitly discussed in Jensen and Meckling (1976), but it truly is the same underlying problem as asset substitution. Debt overhang was the focus of Myers (1977), another extremely famous paper that we unfortunately won't cover in detail.

The general conclusion is that debt induces a preference for risk, and thus distorts incentives about *which projects* to select.

These are deep effects that appear in many models throughout the finance literature, even before Jensen and Meckling (1976), but they especially received attention from the mid-1970s onward. This timing is because options pricing theory, which was rapidly developing around the same time, provides an excellent intuitive framework to understand both effects, as I will discuss below.

As I have tried to emphasize, asset substitution and debt overhang are deeply intertwined. This section will illustrate them formally and simultaneously in a simple framework based on Berkovitch and Kim (1990). At the end of the chapter, I will give a bit more explanation of the connections with options pricing, which are not so clear in this framework.

4.3.1 Analysis following Berkovitch and Kim (1990)

The presentation below closely follows Section 1 of Berkovitch and Kim (1990), which presents a framework that captures both the asset substitution problem discussed in (for example) Jensen and Meckling (1976), and the debt overhang problem discussed in (for example) Myers (1977).

[2] A very similar intuition appeared in Stiglitz and Weiss (1981) in Chapter 2, where the use of a debt contract caused the manager with the riskiest projects to have the strongest desire to invest, even though all projects had the same expected value.

Model setup

- Everyone is risk-neutral and there is no discounting between dates.
- There are three dates:
  - At $t = 0$ the firm issues debt and uses the proceeds to purchase assets, which we refer to as assets-in-place and label X. The debt issued at this date requires the repayment of face value F_0 at $t = 2$. The assets-in-place generate a cash flow at $t = 2$ that depends on the state of the world realized at that date (see below).
  - At $t = 1$ the firm receives a new investment opportunity. It requires investment of I_1, and generates a cash flow at $t = 2$ that depends on the state of the world realized at that date (see below).

    The key focus of our analysis is on whether the firm takes this opportunity, and how closely the firm's decision does or does not align with the socially-optimal decision rule, which is to invest in projects if and only if they have positive NPV.
  - At $t = 2$ the firm realizes cash flows, distributes them to investors, and ceases to exist. There are two potential states of the world at this date, L and H, which happen with probabilities P and $1 - P$ respectively.[3] The assets-in-place generate cash flows of X_L and X_H in these two states. The new project (if it was taken) generates a cash flow of Y in state L, or $Y + s$ in state H. We will consider different values of Y and s, including negative values, to characterize the types of project the firm will and won't pursue at $t = 1$.
- Assume that the investment at date 1 is financed by new debt that is subordinate to the old debt. What this means precisely is the following: In exchange for providing I_1 at $t = 1$, the firm promises to repay face value F_1 at $t = 2$. But the only cash flows available to pay that face value will be whatever is left after repaying the initial lenders their promised amount F_0. The new lenders understand this, and set F_1 accordingly so that they receive enough in expectation to make up for providing I_1.
  - The model's key results are unchanged if the project is financed with equity, or with any security junior to the existing debt. In other words, when we say that this model illustrates the problems with a debt contract, we are referring to the existing debt not the new debt.
  - What we *cannot* assume is that the firm can issue a security senior to existing debt. If this is possible then the problem goes away. This is actually one of the main points made in Berkovitch and Kim (1990).
- Assume that $0 < X_L < F_0 < X_H$, so that the initial debt is risky: without funding the new project, there is a positive probability that the firm cannot make the promised payment F_0.

[3] It's slightly confusing that Berkovitch and Kim choose P to be the probability of the *low* state, but we will follow their notation.

- Assume that $X_L + Y < F_0 + I_1$ so that the new debt would also be risky.

Solution

Note that the NPV of the new project at $t = 1$ is

$$Y + (1 - P)s - I_1 \tag{4.1}$$

In the absence of new investment, the value of existing debt is

$$D_x = PX_L + (1 - P)F_0 \tag{4.2}$$

and the value of existing equity is

$$E_x = \mathbb{E}[X] - D_x = (1 - P)(X_H - F_0) \tag{4.3}$$

The face value of new debt (if issued) must be the solution to

$$I_1 = (1 - P)F_1 + P \times \max\{0, X_L + Y - F_0\} \tag{4.4}$$

And, the market value of equity (if the new investment is funded) is

$$E_1 \equiv (1 - P)(X_H + Y + s - F_0 - F_1) \tag{4.5}$$

We conclude that the firm invests at $t = 1$ if $E_1 - E_x > 0$, where

$$E_1 - E_x \quad = \quad \underbrace{Y + (1 - P)s - I_1}_{\text{Project NPV}} \quad + \quad \underbrace{P \times \max\{-Y, -(F_0 - X_L)\}}_{\text{Distortion}} \tag{4.6}$$

Intuition for the above: The new investment increases equity value by the social NPV of the project, plus an extra distortion reflecting shareholders' desire for risk. Since this extra term does not reflect project NPV, it is completely a transfer from the debt investors. The sign of this term governs the direction of the distortion: If positive, the firm might take an investment that is negative-NPV. If negative, the firm might forgo investment that is positive-NPV.

The result is summarized in their Proposition 1, which we can state here as:

Proposition 4.1. *Let $S \equiv (1 - P)s - I$.*

- *If $S > 0$ and $\frac{-S}{1-P} < Y < -S$, the project is negative-NPV but is funded.*
- *If $S < 0$ and $-S < Y < \frac{-S}{1-P}$, the project is positive-NPV but is ignored.*

In general, equityholders want risk, and this can possibly be more important to them than the actual NPV of the project. Figure 2 of Berkovitch and Kim (1990) illustrate this by plotting a (Y, S) plane, and dividing the plane into regions of positive- and negative-NPV projects, and regions of projects that are ignored and accepted. This figure is reproduced here as Figure 4.1.

Why do these effects happen? While they are mirror images, they sound slightly different when put into words.

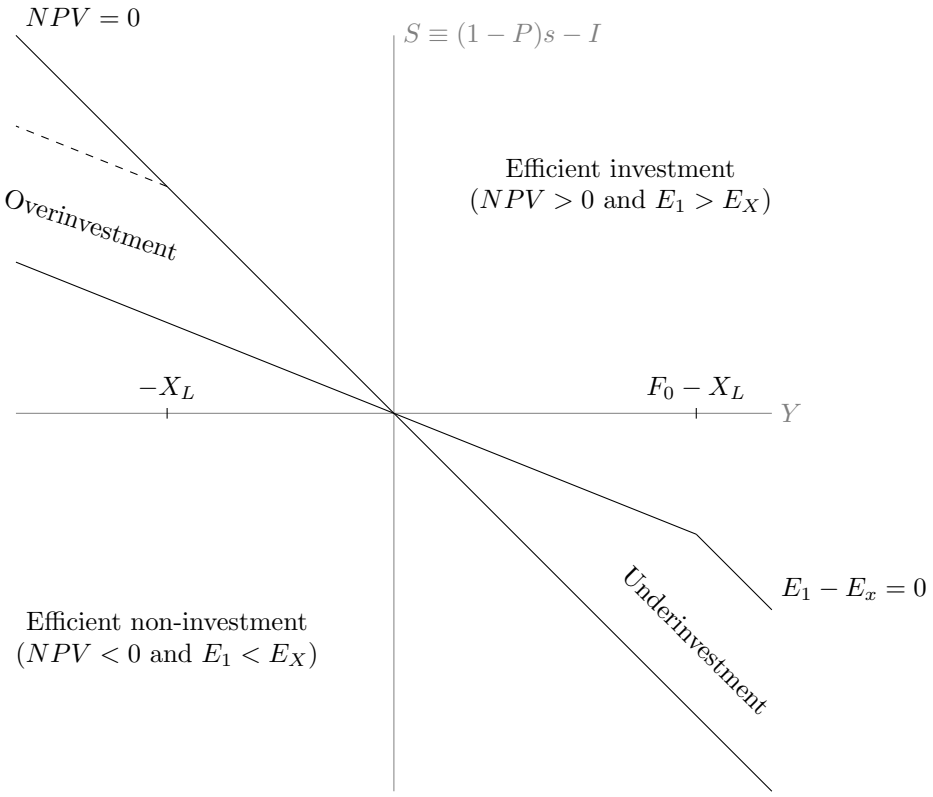


Figure 4.1: Reproduction of Figure 2 from Berkovitch and Kim (1990).

Overinvestment in risky projects (asset substitution) happens when the firm has a large debt burden due in the future, a real probability that it will be unable to make that payment, and the opportunity to gamble on a high-risk project that has *some* probability of paying off enough to save the firm. It will not matter to the firm if that probability is very low and the project is fundamentally negative-NPV: Equity investors will be getting nothing in any case, in any state where the project fails. So all that matters to them is the upside potential of the project.

This situation is depicted in the upper-left region of the figure: These are projects where the certain component Y is very negative, and the possible extra payoff s in state H is positive but not enough to result in positive-NPV. A levered firm may take these projects, against social value, because in any state where s is not realized, the firm is bankrupt anyway.

Underinvestment in safe projects (debt overhang) happens when the firm has a large debt burden due in the future, a real probability that it will be unable to make that payment, and the opportunity to make an investment that will pay off in states of the world where the firm is likely to be bankrupt, but will not pay enough to rescue the firm. In these states of the world, the payoff of the new project will mainly go to benefit old lenders, not equity investors or new lenders. To raise financing to invest in the project, the firm will then have to promise the new lenders cash flows from the good state of the world, and the end result is that equity investors are actually worse off. Hence firms are likely to ignore these projects.

This situation is depicted in the bottom-right region of the figure: These are projects where the certain component Y is positive, but the extra payoff s in state H is relatively small ($S < 0$ means $s < \frac{I_1}{1-P}$). The value of the project is mainly realized in state L where the firm's assets-in-place are doing badly and it is likely to default. These are valuable projects, but the firm will ignore them.

The figure also includes a bit more richness: the lines change slope at a few points. Let's take a few minutes to understand in detail what is happening, as it develops better intuition for the issues we're highlighting. Figure 4.2 modifies the figure for various values F_0 to show how these patterns emerge.

- Start with the case where $F_0 = 0$, which shuts down completely the debt-induced effects that we are talking about. Visually this means sliding the point on the horizontal axis labeled $F_0 - X_L$ all the way over to the point labeled $-X_L$. This is depicted in the top-left panel of Figure 4.2.

  In this scenario, the two lines coincide, except at the top-left, where the $E_1 = E_x$ line will follow the dashed line up and to the left with slope $1 - P$. This reflects an "overinvestment" incentive that is *not* the paper's main focus and arises only due to limited liability (a transfer from society, not from the debtholders). See p.772 of Berkovitch and Kim (1990).

- As we increase $F_0 > 0$, that is, as we begin to slide $F_0 - X_L$ to the right, the inflection point slides down the $-45°$ line, and we see larger amounts

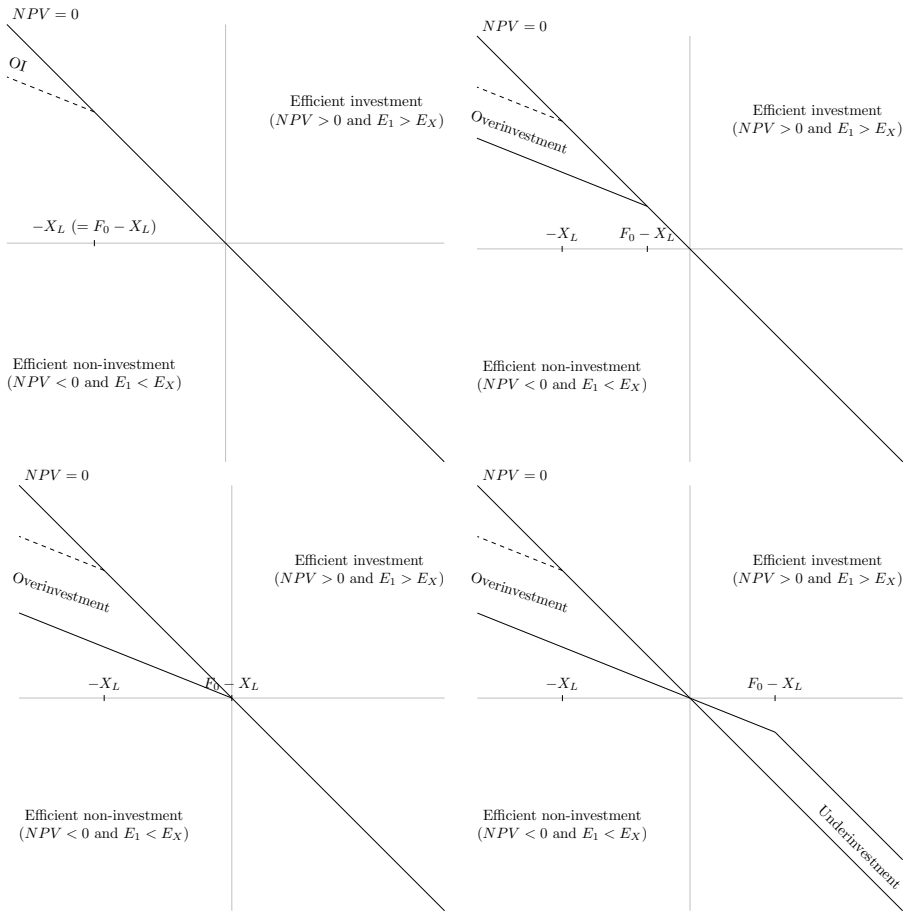


Figure 4.2: Figure 2 from Berkovitch and Kim (1990), modified to have different values of F_0. In the top-left panel $F_0 = 0$, in the top-right $0 < F_0 < X_L$, in the bottom-left $F_0 = X_L$, in the bottom-right $F_0 > X_L$.

of overinvestment, but not yet underinvestment. This is depicted in the top-right panel of Figure 4.2.

- Eventually we slide the point to where it crosses the vertical axis, i.e. $F_0 = X_L$. This is depicted in the bottom-left panel of Figure 4.2. At this point the firm has a large amount of senior debt, but that debt is still risk-free. The overinvestment region is now as big as it will ever get, but we do not yet see underinvestment. The latter problem does not arise until the senior debt is risky.

- Finally as we increase to $F_0 > X_L$, so the point $F_0 - X_L$ crosses the vertical axis, we start to see the underinvestment region develop. This is depicted in the bottom-right panel of Figure 4.2. The point of slope change moves down and right at a slope $1 - P$, causing the underinvestment region to get steadily larger. However, the overinvestment region does not change.

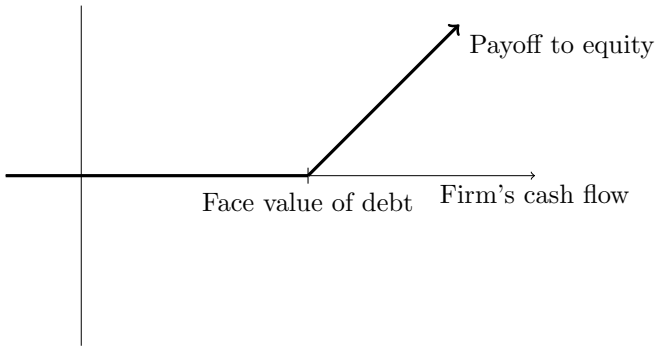
  Intuitively, what matters for overinvestment is the expected recovery value for equity in the *bad* state. As that shrinks, the overinvestment problem grows, but it can't shrink below zero so the problem doesn't get any worse past $F_0 = X_L$. On the other hand, what matters for underinvestment is the difference in the amount paid to old lenders between good and bad states. That difference is zero when the debt is risk-free, so underinvestment does not even happen.

4.3.2 Connections with options pricing

We have illustrated how leverage distorts investment incentives, in a very simple framework with only two possible states, etc. While this framework is tractable, the best *economic* intuition for the problem really comes from drawing a connection to models of options pricing. I will just sketch out the idea here.

We start with the observation that when the firm has debt in place that must be paid before equity, the payoff to equity investors as a function of the firm's overall cash flow resembles the payoff to a call option: If the firm's cash flow is a random variable $\tilde{C}$, and the firm must pay off existing debt with face value F before paying any dividends to equity, then the equity payoff is the random variable $\max\{\tilde{C} - F, 0\}$, which looks like the payoff at expiration of a call option with strike price F and underlying asset $\tilde{C}$. The literature hence refers to the equity of a levered firm as one example of a *real option*.

A basic fact in options pricing theory is that the value of a call option increases with the volatility of the underlying asset's cash flows. Hence, when there is debt in the firm's capital structure, equity investors will have an inherent preference for actions that increase firm risk, completely separate from how those actions affect firm average value (NPV). We can easily set up situations where an action *decreases* firm average value (has negative NPV) but equityholders favor it anyway. In general, the choices of a levered firm will always be distorted somewhat toward riskier actions. This insight generates exactly the agency costs of debt that we saw earlier: A firm with a large amount of debt in place might



take negative-NPV projects if they are very *high* risk (asset substitution), or might ignore positive-NPV projects if they are *low* risk (debt overhang).

The real-options intuition for why leverage distorts the firm's investment policy has been understood since before there were even closed-form solutions for options prices. Myers (1977) repeatedly uses this intuition to explain his result, including the phrase "real option," even though his model is much simpler than the typical options-pricing framework. Following the contribution of Black and Scholes (1973), there was a tremendous amount of work making this analogy more explicit and quantitative. Some prominent papers in this area include:

- Leland (1998) extends the Merton (1974) "structural" (options-based) model of corporate debt valuation, also building on Brennan and Schwartz (1984), to include an asset substitution problem (allowing the firm to adjust the volatility of its cash flow), as well as many other features.
- Hart and Moore (1995) consider a model of capital structure and debt maturity that reflects some of the insights here.
- Lamont (1995) considers whether debt overhang can explain macroeconomic fluctuations.
- Parrino and Weisbach (1999) try to understand the magnitude of these problems through a simulation exercise.

This is indeed a huge literature, but we cannot do it justice without extensively covering the mathematics of options pricing, so for now I will simply leave you with the references above and can provide more on request.

4.3.3 Summary

The agency-based costs of debt described above are a leading candidate for why firms don't use more debt in practice. The idea is that lenders know that the firm's incentives will be heavily distorted as its debt level grows, towards actions that will greatly harm the lender. Hence, they refuse to provide credit beyond a certain point, or demand such exorbitant rates or rigid restrictions that the firm turns to equity markets instead.

Everyone agrees that these effects are real, matter at the margin in some cases, and explain some of what we see in practice. What is still unclear is just how big they are quantitatively, and in particular, whether they are big enough to explain ongoing puzzles, such as why firms don't use more debt, or why investment is slow to recover after recessions.

Some papers that have investigated this question are skeptical. They suggest that these issues only start to bind at extremely high levels of leverage and risk, and simply aren't big enough in practice to explain some of these puzzles. However, there are always methodological debates over how to conduct any such exercise, so we cannot say that the evidence is conclusive. To investigate this thoroughly would be a topic for an empirical course.

In any case, measuring the magnitude of agency problems at the levels of debt that we *typically observe in practice* is not necessarily what we want to do anyway. The argument of Jensen and Meckling (1976) is that agency problems are the reason firms don't have *much higher* leverage ratios than what we typically observe, like 90% or more. It is difficult to test this argument empirically, precisely because such leverage ratios are so far outside the normal range of data, which may itself be a rational decision to avoid the very agency problems that we are talking about.

4.4 Conclusion

We have demonstrated that both equity and debt contracts can distort the firm's investment policy due to problems of incentives. The problem with equity financing is that it significantly weakens the incentives of the firm's insiders, as their ownership stake falls. The problem with debt financing is that it skews the firm's preference towards investments with high risk, potentially sacrificing NPV and creating transfers from lenders to shareholders.

Section 5 of Jensen and Meckling (1976) suggests that these insights can be drawn together into a unified "theory" of capital structure based on agency problems. They basically propose that equity and debt in the capital structure generate agency costs in the respective manner described above, and the firm chooses capital structure to minimize the sum of those costs. It much like the tradeoff theory, but the costs and benefits are those outlined in this chapter. This "theory" has been quite influential in the long literature building on Jensen and Meckling (1976) and investigating these issues empirically.

However, this being a theory course, we should note that they only sketch this final section intuitively. They never rigorously analyze a model that combines *both* types of agency problem (effort choice and project choice) along with *both* types of contract (equity and debt) being available to finance investment. Hellwig (2009) attempts this analysis, and finds it is nowhere near as straightforward as suggested by Jensen and Meckling (1976).

We will not attempt cover Hellwig's analysis in detail, but it highlights an important observation for us. When you try to combine theoretical insights from separate, restricted models into a single framework, they often interact

in ways that surprise you and lead to conclusions that are ambiguous or even counter to what you expect. This is why theory cannot be done verbally!

Ultimately, the legacy of Jensen and Meckling (1976) is less in its theoretical framework, and more in the way it convinced many people to take incentives issues seriously as a first-order explanation of stylized facts in corporate finance and investment.

4.5 Exam question: Agency problems of debt

4.5.1 Debt overhang

Consider the following model based on Berkovitch and Kim (1990):

- Everyone is risk-neutral and there is no discounting between dates.
- A firm has debt with face value of F_0 due at time 2.
  The firm's equity investors receive any residual cash above F_0 at time 2.
  The firm acts to maximize the expected value of this equity payoff.
- The firm's existing operations generate a random cash flow at time 2:
  With probability P, the state is H, and the operating cash flow is X.
  With probability $1 - P$, the state is L, and the cash flow is zero.
- At time 1, the firm also has the opportunity to invest in a new project.
  The project requires investment of I at time 1, and generates a cash flow of Y at time 2 (regardless of the state). Assume that $I < Y < F_0 < X$.
- The new project, if taken, must be financed by new debt, consisting of a face value F_1 due at time 2, that is subordinate to the old debt.

Do the following:

1. If the firm decides to invest at time 1, what face value F_1 must it promise?
2. Based on this, derive a condition on P under which the firm's decision at time 1 will be inefficient. Give some intuition why this problem arises.
3. Suppose that the firm can issue *senior* debt at date 1. That is, new debt would have first claim on cash flows at time 2, then old debt, then equity. How would this change your analysis from the prior question, and why?

4.5.2 Asset substitution

Continue the setup from the previous section, but make the following changes:

- Assume that the project generates the positive cash flow Y only in state H, while in state L the project will generate a *negative* cash flow of $-Y$.
- Assume that the firm has cash equal to I initially. Hence, the firm no longer needs to raise external financing to invest. The firm's cash can be invested in the project, or can be saved until date 2 at zero interest.

Do the following:

1. Derive a condition on P under which the firm's decision at time 1 is inefficient. What is different about this from the prior section?
2. Now assume the firm could also pay out the cash on hand as a dividend to equity investors at time 1. Under what condition will the firm choose to do this? How much would the existing lenders be willing to pay, to add a covenant in the debt contract that forces the firm to save its cash?

Solution to exam question 4.5

Debt overhang

1. If the firm invests, in state H it will have $X + Y - F_0 > 0$ available to repay the new debt. In state L it will have nothing available, as its only cash flow will be X and this will be entirely taken by the old lenders since $X < F_0$. So the face value F_1 that the firm promises can only be paid in state H, so it must satisfy $PF_1 = I$ or in other words $F_1 = I/P$.

   Note that it is not trivial that the firm will even have that debt capacity: If $P \geq \frac{I}{X+Y-F_0}$ then the firm can promise $F_1 = I/P$ to fund the investment. But if $P < \frac{I}{X+Y-F_0}$ then there is no face value F_1 that the firm can promise that will raise the necessary financing of I, so investment simply can't happen.

2. The efficient decision is to fund the new project since $Y > I$ (and there is no discounting). So we are looking for a condition under which the firm will not do this. We already have one sufficient condition from the prior question: $P < \frac{I}{X+Y-F_0}$. But there is a weaker sufficient condition: Even if the firm can raise the necessary debt financing, it might choose not to.

   To spell this out explicitly, if the firm does not invest, its equity is worth $P(X - F_0)$. If it does invest by issuing new debt with face value as derived in the prior question, its equity will be worth $P(X + Y - F_0 - F_1) = P(X + Y - F_0) - I$. It will choose to ignore the investment iff $P < I/Y$. This is a strictly weaker condition than $P < \frac{I}{X+Y-F_0}$.

   Intuitively, even when the project is efficient and the firm can finance it, investment may not be in the interests of equity holders. They will ignore the payoff of the project in the state of the world where the firm is insolvent, because this does not benefit them, and focus only on the payoff in the state of the world where the project can benefit them at the margin. Hence they ignore the social expected payoff of Y and consider only their own expected payoff PY. As the firm is at greater risk of insolvency (as P falls), it becomes more difficult to convince them to invest. This is the essence of a *debt overhang* problem.

3. If the firm can issue senior debt, then it will always have capacity to pay that debt in either state. In the previous question, in state L the new cash flow Y from the new investment had to go to service old debt, but in this new setup that cash flow can service new debt. Then, the face value of the new debt only needs to be $F_1 = I$ as it is now risk-free. The firm will now invest iff $P(X + Y - F_0 - I) > P(X - F_0)$, and this is guaranteed by the assumption $Y > I$.

   In words, when new lenders are senior, they can extract their repayment in both states of the world, hence demand a lower face value and leave more for equity in the good state. Equity investors now internalize the social value of the project.

Asset substitution

1. If the firm does not take the new project, its equity value is $P(X+I-F_0)$. If the firm does take the new project, the equity value is $P(X+Y-F_0)$. Since $Y > I$, the firm will take the new project. This is inefficient from a social perspective if $P < 1/2$.

   The similarity to the previous section is that equity investors ignore the state of the world in which the firm is insolvent. The difference is that in the prior section, equity investors ignored a state with positive payoff, and hence ignored a valuable project. Here they ignore a state with negative payoff, and hence accept an inefficient project.

2. If the firm pays out cash, its equity value is $P(X-F_0)+I$. This is greater than $P(X+I-F_0)$ so the firm will never retain the cash. Payout is better than investment if $I > PY$, which happens to be the same condition as we derived in the prior section – this is not surprising, since the payout decision is the mirror image of the decision to finance with junior debt.

   In either case, lenders lose PI of value compared to if the firm retained the cash on the balance sheet, so they would be willing to pay up to this amount for the covenant described in the question.

4.6 Exam question: Culture and compensation

1. Suppose a firm will generate profits of $e\,Y$, where e is an effort level chosen by the manager. Suppose that effort carries a convex disutility cost $c(e)$. Finally, suppose that the manager is risk-neutral, and is compensated with a fraction α of the firm's profits, so her payoff for any effort choice e is $\alpha eY - c(e)$. Show that if the firm's insiders wanted to maximize profits $e\,Y$, then they would have to set $\alpha = 1$.

2. Now we reinterpret the above situation slightly:

   Suppose the manager can allocate effort between improving the firm's profits, and improving its culture, with total effort across these two tasks adding up to a fixed E. Suppose that the manager's utility from allocating effort e towards profits, and $E - e$ to culture, while receiving the fraction α of profits as before, is given by $\alpha eY + \ln(E - e)$. This provides a specific functional form for the disutility cost $c(e)$ from the prior question.

   If the firm's insiders just wanted to maximize $e\,Y$, then we would know from the prior question that they have to set $\alpha = 1$. But suppose we instead assume that the firm's insiders also care about culture. Specifically, suppose insiders want to maximize $eY + \ln(E - e - \kappa)$, where $\kappa > 0$.

   However, suppose that the insiders also cannot directly measure the firm's culture or the manager's effort in improving it. So they simply continue to compensate her with a fraction α of profits.

   In this case, show that insiders now optimally set α to be *less* than 1. Give some interpretation of this result.

3. Finally, imagine that investors in the stock market also care about the company's culture, and this affects their valuation of the stock, separately from the company's profits. What opportunity does this create to improve on the compensation contract we have analyzed so far?

   (You don't need to show any formal analysis, just describe what new feature we could add to the contract that would improve things from the perspective of the insiders.)

Solution to exam question 4.6

1. The manager's chosen effort level e^* will solve $\max_e \alpha eY - c(e)$, which means it must satisfy the first-order condition $\alpha Y = c'(e^*)$. Since c is convex in e, then c' is increasing in e, so e^* is increasing in α. Then if the insiders' goal is to maximize output eY, this means maximizing e^*, which means maximizing α. Since α cannot be any higher than 1, to maximize output they would have to set $\alpha = 1$.

2. Again the manager's chosen effort level e^* will satisfy $\alpha Y = c'(e^*)$, but here we have a specific value $c(e) = -\ln(E - e)$, so $c'(e) = \frac{1}{E-e}$, and we can then solve the FOC directly to get $e^*(\alpha) = E - \frac{1}{\alpha Y}$.

   Meanwhile, the insiders' maximization problem can be described as

   $$\max_{\alpha} e^*(\alpha)Y + \ln(E - e^*(\alpha) - \kappa)$$

   which generates their own first-order condition

   $$(e^*)'(\alpha)Y = \frac{1}{E - e^*(\alpha) - \kappa} \times (e^*)'(\alpha)$$

   and we can substitute our earlier expression for e^*, and solve, to get

   $$\alpha = \frac{1}{1 + Y\kappa}$$

   Since $Y > 0$ and $\kappa > 0$, we conclude that the insiders' optimal choice of α is now less than 1.

   Interpretation: Now the firm's insiders care about the firm's culture separately from its sales. In a perfect world, they would compensate the manager directly for her effort on both sales and culture, designing the contract to reflect their own attitude about the tradeoff between these two. But by assumption they have no way to directly accomplish this. On the other hand, the manager has an innate tendency to invest in culture when not distracted by sales-focused compensation, because she cares about it too. Then it can make sense to ease off on the manager's sales compensation, to give her the slack to invest in culture.

3. In this case, the natural thing to do is to provide the manager with compensation tied to the company's stock price. This could mean grants of stocks, options, or warrants, or bonuses tied to the share price. The exact weights that the optimal contract would place on any of these as opposed to profits or other measures would depend on the details of how much the market valued culture. The bigger point is that focusing on shareholder value does not inherently mean ignoring factors other than profits, if the investors in the market also care about those other factors.

Chapter 5

Hidden action and security design

5.1 Overview

In the prior chapters, we've pointed out various costs and benefits of debt and equity contracts, the most common financial contracts that firms use. At some point, the natural question to ask is why these contracts should be the standard in the first place.

The question is particularly salient for debt contracts. Equity contracts at least have an intuitive structure: Cash flows are shared roughly according to dollars invested. Indeed, one can almost imagine that this should have been the benchmark. By contrast, debt has a structure that seems simple at first (just make a fixed payment), but is far more complex once we take into account that the borrower may be unable to do that. The contract effectively awards every marginal dollar to the lender up to a threshold, then every marginal dollar to the borrower after that. Why is it so popular to use a contract that induces such a stark change in marginal payoffs and incentives around a single value of the cash flow realization?

These questions bring us to the topic of security design. This literature attempts to derive the *optimal* contract as a maximization problem, taking as given the economic environment, including the parties that are to sign a contract, the frictions they face, and any constraints on the contract space.

We will focus only on security design with hidden actions (which is a subset of the much broader literature on optimal contracting). There is also a literature on security design with hidden information (e.g. Nachman and Noe, 1994, which we mentioned in passing in Chapter 3), but this literature is smaller and we will not have time to cover it. Even within this focus, the literature includes a wide range of important questions and contributions, with examples including Chiesa (1992), Chang (1993), and Repullo and Suarez (1998).

We will focus on just two famous models: first, Innes (1990), and second,

the "costly state verification" model (really a class of models spanning several papers). Both are simple settings that also illustrate some important technical points for us. And, they share a similar message: In each paper, a reasonable set of frictions and contract restrictions leads to a simple debt contract emerging as the optimal contract.

This chapter draws heavily on Chapter 3 of *The Theory of Corporate Finance* (Tirole, 2006), and you may find it useful to read the discussion there as well.

5.2 Innes (1990)

Recall that Jensen and Meckling (1976) considered a problem with a hidden choice of effort level, and argued that firms should therefore avoid equity financing. They imply that this should push the firm to use debt instead, assuming that these are the only two available choices. But could there be some other contract that is better than either debt or equity for solving this problem? Innes (1990) shows that debt can actually be derived as the *optimal* contract in this problem, under certain assumptions.

We can start with a couple of standard benchmark observations. With risk aversion and no hidden action, the entrepreneur should just "sell the project" and work for a fixed wage. In reverse, with risk neutrality, hidden action, and *unlimited* liability, the entrepreneur should offer the principal a fixed repayment and consume the residual.[1]

Innes starts from the second situation above, and adds limited liability to the borrower. This is the most important ingredient in the analysis. However, it does not yet generate debt as the optimal contract. Instead the optimal contract takes a "live or die" form (described below). To generate a realistic debt contract, Innes (1990) then applies a "monotonicity" constraint based on arguments that feel reasonable but are outside the model. We will consider these two steps of the argument separately.

5.2.1 Model setup

The exposition below closely follows Tirole's textbook, starting on his page 132.

- The manager has assets A to invest in a project that requires funds of I, so, the manager must raise financing of $I - A$ in order to invest.
- The cash flow generated by the project is a random variable R that is distributed on $[0, \bar{R}]$ according to the density function $p(R|e)$. The density is conditioned on e, which represents the manager's choice of effort.
- We assume p satisfies the *monotone likelihood ratio property* (MLRP):

$$\frac{\partial}{\partial R}\frac{\partial p(R|e)/\partial e}{p(R|e)} > 0$$

[1] See for example Shavell (1979) and Harris and Raviv (1979).

In words, this means that an increase in the manager's effort causes a uniform rightward shift in (the log of) the conditional density function $p(R|e)$. Conditions like this are very common in the literature on hidden actions (the classic reference is Milgrom, 1981). In our setting, the MLRP just imposes the minimal notion that the payoff of the project is informative about the effort level of the agent, so rewarding the manager for payoff is an indirect way of rewarding him for effort.

- The manager is risk-neutral. His utility is equal to the expected payoff of the contract given his effort level, minus a disutility of effort g. We assume $g' > 0$, $g'' > 0$, $g(0) = g'(0) = 0$, $\lim_{x\to\infty} g'(x) = \infty$. The first two conditions are important, the rest are mainly for convenience.

- The contract between manager and investor specifies a payoff $w(R)$ to the manager as a function of project realization R. Note that we are assuming you can contract directly on the realization of R.[2]

5.2.2 Formal justification for the "live-or-die contract"

First we show why the "live-or-die" contract emerges from Innes's model when we don't impose a monotonicity constraint on the contract. Without the monotonicity constraint, the entrepreneur's problem is

$$\max_{w(\cdot),e} \int_0^{\overline{R}} w(R)p(R|e)dR - g(e)$$

subject to the constraints

$$\int_0^{\overline{R}} w(R)\frac{\partial p(R|e)}{\partial e}dR = g'(e) \qquad \text{(IC for borrower)}$$

$$\int_0^{\overline{R}} [R - w(R)]p(R|e)dR \geq I - A \qquad \text{(IR for lender)}$$

$$0 \leq w(R) \leq R \qquad \text{(limited liability, feasibility)}$$

- The first line above is the manager's objective function.

- The second line above is an *incentive compatibility* constraint: The choice of effort e must be optimal ex-post from the manager's perspective. This equation is just the first-order condition of the manager's problem to maximize e taking as given the contract $w(\cdot)$, which is exactly the problem that the manager will solve after the contract is signed.

- The third line above is an *individual rationality* constraint (sometimes called a *participation* constraint): For it to be rational for the lender to provide funds to the project, she must expect to receive at least as

[2]Note also a cosmetic difference from the original Innes paper: He defines the contract as specifying not the borrower's payoff w but rather the lender's payoff $B(R) \equiv R - w(R)$.

much value as the funds being provided. The overall problem maximizes borrower utility while giving the lender just enough payoff to participate. Thus, as with many of the models we studied in earlier chapters, we are implicitly imagining a fairly competitive lending market in which all profits are competed away.

- In the fourth line, the first half of the inequality $w \geq 0$ assumes that you cannot punish the manager beyond taking the project cash flow away. Innes calls this a "limited liability" constraint. It is not critical that the lower bound here is zero, but, it is important that there is a lower bound.[3] The second half of the inequality assumes that entrepreneur cannot be rewarded with anything more than the cash flows from the project itself. This can be motivated as an assumption that any greater bonus would be infeasible, or would violate some kind of limited liability by the investor.

Write out the problem in Lagrangian form, attaching multipliers μ and λ to the borrower and lender IC constraints:

$$\begin{aligned}\mathscr{L} \equiv & \int_0^{\overline{R}} w(R)p(R|e)dR - g(e) + \mu\left[\int_0^{\overline{R}} w(R)\frac{\partial p(R|e)}{\partial e}dR - g'(e)\right] \\ & + \lambda\left[\int_0^{\overline{R}} [R - w(R)]p(R|e)dR - I - A\right] \qquad (5.1) \\ = & \int_0^{\overline{R}} w(R)p(R|e)\left[1 + \mu \times \left(\frac{\partial p(R|e)/\partial e}{p(R|e)}\right) - \lambda\right]dR \\ & + \quad \lambda\int_0^{\overline{R}} Rp(R|e)dR - I - A - g(e) - \mu g'(e) \qquad (5.2)\end{aligned}$$

In (5.2), note that all the terms that depend on w appear inside the first integral. Everything after this integral is irrelevant to characterizing optimal w. Moreover, we can see that the optimal contract will seek out "corner solutions." For values of R where the bracketed expression is positive, the optimal contract will set w as high as possible, and where the expression is negative, it will set w as low as possible. Imposing our condition $0 \leq w(R) \leq R$ we conclude that

$$w(R) = \begin{cases} R \text{ if } 1 + \mu\frac{\partial p(R|e)/\partial e}{p(R|e)} > \lambda, \\ 0 \text{ if } 1 + \mu\frac{\partial p(R|e)/\partial e}{p(R|e)} < \lambda. \end{cases} \qquad (5.3)$$

[3]Without this assumption, the optimal solution would be for the borrower to simply promise a completely risk-free payment to the lender in every state of project realization, and accept all project risk himself. This is a standard result in hidden-action models with risk-neutral agents and is often described as the borrower "buying the project" or "buying the output." Outside corporate finance, this result can be overturned by assuming the borrower is risk-averse, in which case "buying the project" would expose him to too much risk to be optimal. Innes wants to stick with risk-neutral agents, which is more standard in finance, so it is the limited-liability constraint and not risk aversion that prevents this outcome.

We restrict attention to problems where μ is strictly positive, i.e., the effort level that is most attractive for the manager ex post is not the one he would commit to ex ante, if he could. Then, the MLRP assumption implies that there is some cutoff R^* that separates the two regions above, and the optimal contract gives the manager zero payoff for realizations below R^*, and the entire project value above this. From the lender's perspective, the optimal contract pays the entire project value for realizations below R^*, and nothing above this.

This is what Innes calls a "live-or-die" contract. It provides maximal incentives to the entrepreneur, conditional on providing sufficient payment to the lender to break even. But the odd thing about it (compared to the types of contracts we see in practice) is that it gives *everything* to the entrepreneur provided only that the project payoff is above some lower bound, and everything to the lender otherwise. As project return crosses the lower bound, payoffs shift dramatically away from the lender and towards the borrower.

This is the correct answer to the problem that we posed, but it feels unrealistic. Next Innes considers what reasonable restrictions in practice might cause the optimal contract to look more realistic, and perhaps more like debt.

5.2.3 Heuristic argument for debt with monotonicity

Innes next imposes the monotonicity constraint:

$$R - w(R) \text{ nondecreasing in R} \qquad (\text{ "monotonicity"})$$

The idea is that it is simply unrealistic to specify a contract that sharply shifts payoffs away from the lender as project cash flow marginally increases, as was key in the live-or-die contract. Innes provides an intuitive argument that if the manager has some ability to conceal the source of cash flows, then if the project payoff is just below the discontinuity in the contract, he can always scrounge up some financing from another source to push him over the edge and report R^* as the project's cash flow.

It is always a little uncomfortable to reason loosely in this manner about things that could be modeled explicitly. (And indeed we will model explicitly the possibility that the manager misreports cash flows, with the "costly state verification" models in the next section.) However, the overall idea does sound reasonable. As with the intuitive criterion or other equilibrium refinements we have imposed, sometimes it is worth seeing what happens when we impose "natural" conditions on a problem, without being too rigorous about exactly how those conditions would arise.

Innes shows that with the monotonicity constraint, and a few further technical assumptions, the problem yields debt as the optimal contract. We will not go through the proof as it requires many uninteresting details. (The strategy is largely to prove the result bit-by-bit by contradiction.) But the intuition behind the proof can be illustrated with just the following heuristic argument, which is also quite similar to a proof of a result in the next section:

Suppose there was a monotonic *non-debt* contract w^{ND} that solved the problem. First, you show that it's possible to construct a debt contract w^D that yields the same expected payoff to investors, *holding fixed* the manager's effort level at that induced by the non-debt contract. This is just a matter of reallocating payoffs appropriately across realizations R.

Then, you can show that when the manager is allowed to choose effort freely, *the debt contract actually induces a higher effort choice than the non-debt contract.* This is the key to the entire result. By construction, lenders and borrowers receive the same expected payoff under the old and new contract, at the borrower's old effort choice. By assumption, the old contract was not debt, but is monotonic and respects the feasibility constraint $0 < w^{ND}(R) < R$. We can show that $w^{ND}(R)$ must cross $w^D(R)$ from above at some point (or set of points), that is, $w^{ND} > w^D$ for sufficiently low R, $w^{ND} > w^D$ for an intermediate range of R (possibly degenerate), and $w^{ND} < w^D$ for sufficiently high R. In a nutshell, the debt contract offers payoffs that are more sensitive to project returns. Given this, the MLRP assumption will imply that the manager optimally exerts more effort under the debt contract.

Because the debt contract induces higher effort, it also creates strictly more surplus that can be shared among everyone, compared to the non-debt contract. From this we can build a debt contract that dominates the original non-debt contract: Since everyone is risk-neutral, we can adjust the face value on the debt contract to reallocate surplus between the loan parties until the lender again breaks even, and then the borrower must be getting strictly more surplus than under the old contract. This contradicts the assumption that the non-debt contract solved the problem in the first place.

5.2.4 Discussion

The model of Innes (1990) focuses entirely on the borrower's effort incentives, very similar to the first section of Jensen and Meckling (1976). It formalizes the notion that you want a contract that preserves the borrower's payoff being as close to the 45° line as possible, epsecially in good states. For a company run by a single person who has never raised outside finance, their payoff is naturally this way. If the company is forced to raise outside finance, then among monotonic financing contracts, debt is the one that does the least to distort the borrower's incentives away from this ideal.

It's very important to emphasize that the borrower in this model is risk-neutral. As his risk aversion grows, debt becomes a very *bad* contract for him in the sense that it forces lots of risk on him. This is a deep tension between *insurance* and *incentives* that appears throughout the literature on optimal contracting. It's difficult to give people the security of a safe baseline wage without destroying their incentives to exert effort.

5.3 Costly state verification

This literature considers a different type of agency problem: Now the distribution of cash flows does not depend on any effort or project choice by the manager, but, the manager is the only person who will certainly *observe* those cash flows. Lenders must expend some resources if they want to observe them. In this setting, debt often emerges again as the optimal contract, but here the motivation is to economize on expected monitoring costs.

The presentation below closely follows Tirole (2006) starting on his p.138. The typical citation that people make for the general idea of costly state verification is Townsend (1979), so you should know that reference. However, his setting is actually quite different from what we typically have in mind with finance or accounting models: it assumes a risk-averse borrower, and focuses a lot on issues that are specific to insurance markets, as well as general equilibrium issues. Gale and Hellwig (1985) is a bit closer to the presentation below, and the typical finance/accounting setting, as it assumes risk-neutral agents. Diamond (1984) adapts these ideas into a theory of banking (see next chapter).

5.3.1 Model setup

- Everyone is risk-neutral.
- The borrower has wealth of A and the opportunity to pursue a project that requires investment of I, so must raise funds equal to $I - A$.
- The project's payoff is $R \sim p(\cdot)$, which is *not* affected by anyone's actions.
- After R is realized, the borrower observes it, and sends a report $\hat{R}$ to the investor about it. The investor can also observe R directly if they choose, but this requires paying an "audit cost" $K > 0$.
- Investment contracts map reports $\hat{R}$ to three outcomes:
  - a decision whether to audit $y(\hat{R}) \in \{0, 1\}$;[4]
  - a reward function for the borrower $w_1(\hat{R}, R) \geq 0$ if $y(\hat{R}) = 1$, with the lender getting the residual $L_1(\hat{R}, R) \equiv R - w_1(\hat{R}, R)$;
  - and a separate reward function $w_0(\hat{R}, R) \geq 0$ if $y(\hat{R}) = 0$, with the lender getting the residual $L_0(\hat{R}) \equiv R - w_0(\hat{R}, R)$.

Note that the last item in the definition indirectly imposes structure on w_0 by specifying that the lender's payoff L_0 when not auditing can only be a function of the report, not the true state, which the lender does not observe.

(It might seem clearer to define L_1 and L_0 as the objects of the maximization, and w_1 and w_0 as simply the residual of project payoff over these functions, but I am trying to stay close to Tirole's notation.)

[4] A separate branch of this literature (e.g. Mookherjee and P'ng, 1989, and a section in Townsend, 1979) allows lenders to *randomize* their audit decisions, so we would say $y \in [0, 1]$ represents a probability of audit, not a deterministic choice. Here the optimal contract often is *not* debt. Random audits sound odd, but these papers argue that they are realistic.

5.3.2 Formulating the problem: The revelation principle

We assume that the borrower will design the contract, and will choose it to maximize his own expected payoff, subject to two constraints on the problem.

The first constraint is obvious: investors must expect a return of at least $I - A$. Once again, this is a problem where we maximize the payoff to the borrower while giving the lender just enough to justify their participation, which implicitly imagines a competitive lending market that is not modeled explicitly.

The second constraint we will impose is much less obvious: The borrower must find it optimal to report truthfully given the contract terms. That is, he must expect a weakly higher payoff from setting $\hat{R} = R$ than from any other report. This should sound very odd and it deserves discussion.

First, understand that this is a restriction on the contract, not on the borrower's behavior. That is, we are not directly constraining the entrepreneur to tell the truth. Rather, we are constraining the optimal contract to be one that *incentivizes* the borrower to tell the truth. The thing that we need to explain is why the optimal contract should have this property.

This is due to a very general result in mechanism design known as the **revelation principle**. This result says that when solving for optimal contracts, we "lose nothing" by restricting to contracts that induce a weak preference for truth-telling. Or to put this differently, any allocations and outcomes that can be achieved by a general contract, can also be achieved by some contract that induces truth-telling. So the contract we find after imposing the truth-telling constraint, will be at least weakly superior to any possible contract.[5]

The revelation principle greatly simplifies the process of finding optimal contracts. We only need to find the best truth-telling contract, which we can do by (1) adding the constraint described above to our optimization problem; and (2) otherwise assuming truth-telling in the rest of the problem.

[5]The revelation principle may sound implausible when you first hear it, but the proof is actually obvious once you know it: Take any contract $w(\hat{R})$, which is a mapping from reports about R to cash flows, and consider the borrower's optimal strategy $\sigma_w(R)$ in response to w, which is a mapping from realizations of R to reports about R. We can define a second contract by $(w \circ \sigma_w)(\hat{R})$. That is, whatever cash flow the entrepreneur reports, the contract will translate it to a "report about the report" following the strategy σ_w that was induced by w, and then will translate that new report to cash flows following the original contract w. Or to put it differently, this contract plays the borrower's strategy σ_w for him. If σ_w was optimal behavior in response to w, then truth-telling should be an optimal behavior in response to $w \circ \sigma_w$. It follows that whatever we can accomplish with general contracts, we can accomplish with contracts that induce truth-telling.

It is unclear who exactly deserves credit for the revelation principle. It seems to have been widely known in some form before being written down, and then was codified and extended over the course of several papers. Papers will often list some or all of the following citations: Gibbard (1973), Green and Laffont (1977), Dasgupta et al. (1979), Myerson (1979).

Hence the optimal contract solves

$$\max_{y(\cdot),w_0(\cdot),w_1(\cdot)} \int w(R)p(R)dR \tag{5.4}$$

subject to

$$\int [R - w(R) - y(R)K]p(R)dR \geq I - A \tag{5.5}$$

$$w(R) = \max_{\hat{R}} (1 - y(\hat{R})) \times w_0(\hat{R}, R) + y(\hat{R}) \times w_1(\hat{R}, R) \tag{5.6}$$

(5.4) is the borrower's objective function, (5.5) is the lender's "participation constraint" or "individual rationality constraint" (IR), and, (5.6) is the truth-telling constraint due to the revelation principle. Notice how the first two lines are written with the assumption that the borrower tells the truth.

5.3.3 Solving the problem and deriving the debt contract

Define a debt contract as a contract that specifies, for some number D,

- $y(\hat{R}) = \mathbb{1}\{\hat{R} < D\}$,
- $L_0(\hat{R}) = D$ and hence $w_0(\hat{R}, R) = R - D$,
- $L_1(\hat{R}, R) = R$ and hence $w_1(\hat{R}, R) = 0$.

That is, the borrower pays the "face value" D if possible. If the borrower reports being unable to do so, then the lender audits and receives any available cash.

Now we show that the optimal contract has this structure almost-everywhere:

We can first observe that (5.5) must bind at the solution to the problem (as always in problems like this). Intuitively, since the borrower designs and offers the contract, it would never make sense to offer any more than necessary to the lender. So we can rewrite (5.5) as an equality, and then solve it for the borrower's objective function:

$$\int w(R)p(R)dR = \int [R - y(R)K]p(R)dR - (I - A)$$

We can substitute this into the objective function of the original problem, remove all the terms that do not depend on any of the choice variables, and see that the problem is really just to minimize the probability of audit:

$$\min_{y(\cdot),w_0(\cdot),w_1(\cdot)} \int y(R)p(R)dR \tag{5.7}$$

subject to

$$w(R) = \max_{\hat{R}} (1 - y(\hat{R})) \times w_0(\hat{R}, R) + y(\hat{R}) \times w_1(\hat{R}, R) \tag{5.8}$$

This pattern arises often in economic models: The borrower is not the one who directly pays the audit costs, yet he still internalizes them, in the sense

that the problem of maximizing his payoff is exactly identical to the problem of minimizing the expected amount of audit costs paid. By saving on the lender's audit costs, he can compensate by offering less of the project cash flow to the lender, so ultimately this is to his own benefit.

Because audits are deterministic, contracts must partition the set of reports $\hat{R} \in [0, \infty)$ into an "audit" region and a "no-audit" region. For any given contract, denote these regions by $\mathcal{R}_1 \subseteq \mathbb{R}$ and $\mathcal{R}_0 = \mathbb{R} \setminus \mathcal{R}_1$ respectively.

Next we make a few observations about the optimal debt contract that follow from the truth-telling constraint:

- The payment to the lender must be constant for all reports in $\mathcal{R}_0$: If not, for any $R \in \mathcal{R}_0$ the borrower would strictly prefer to report whatever $\hat{R}$ generates the lowest payment, violating (5.8). So, for all $\hat{R} \in \mathcal{R}_0$, we have $L_0(\hat{R}) = D$ and $w_0(\hat{R}, R) = R - D$ for some D.

- The payment to the lender for *any* report is at most D: If not, so there is some report $R^\circ \in \mathcal{R}_1$ for which the lender receives strictly more than D, the borrower would report some $R' \in \mathcal{R}_0$ when $R = R^\circ$, violating (5.8).

- The lender must audit any report less than D: If not, then there is some $R^* < D$ for which $R^* \in \mathcal{R}_0$ and hence (from above) $L_0(R^*) = D$. Then the contract would have to specify $w_0(R^*, R^*) < 0$, violating $w_0 \geq 0$.

Finally we demonstrate two more key properties that the optimal contract must have almost everywhere. We will do this by showing that for any contract that does *not* feature the property in question, we can replace it with another one that does, and the latter contract will satisfy all constraints on the problem, while generating strictly less expected audit cost.[6]

- Under the optimal contract, for reports greater than that contract's D, the lender chooses no-audit almost everywhere. *Proof:* Take any contract $\mathcal{C}$ that satisfies the constraints on the problem. Suppose there is an interval of reports greater than $\mathcal{C}$'s value of D at which the lender audits. As we have already noted, the lender's payment in this region must be D. So we can replace $\mathcal{C}$ with another contract in which the lender does not audit any report greater than D, and this will generate the same expected payoff to the lender, with a strictly lower probability of audit. This new contract satisfies all the constraints on the problem and is strictly preferred to $\mathcal{C}$, so $\mathcal{C}$ cannot have been the optimal contract.

- Under the optimal contract, for reports less than that contract's D, the lender's specified payment almost everywhere is the entire project value R. *Proof:* Take any contract $\mathcal{C}$ that satisfies the constraints on the problem. Recall that we have already shown the lender will always audit any report less than $\mathcal{C}$'s value of D. Suppose there is an interval of such reports for

[6]Because the objective is based on *expected* audit costs, this strategy cannot rule out anything about a contract's behavior on sets of zero probability measure. That is why this last section and the model's overall conclusions are only true "almost everywhere."

which the lender receives less than the full project cash flow. We can replace $\mathcal{C}$ with another contract in which the lender receives the entire project cash flow for these reports, while lowering the contrract's value of D just enough to match the lender's expected payoff under $\mathcal{C}$. Since it has a strictly lower value of D, the new contract has a strictly lower audit probability than $\mathcal{C}$. So it satisfies all the constraints on the problem and is strictly preferred to $\mathcal{C}$, so $\mathcal{C}$ cannot have been the optimal contract.

We conclude that the optimal contract is a debt contract almost everywhere.

5.3.4 Discussion

The proof in the prior section is a bit tedious, but the intuition is extremely clear: It seems natural to suppose that lenders face greater costs than borrowers in verifying (auditing) how much cash is available to repay their investments. Then they would clearly like to minimize the amount of this auditing. But, that also creates obvious incentives for borrowers to always report the lowest cash flow that will not trigger an audit. A standard debt contract provides the lender with their required rate of return at the lowest possible audit cost: As long as the firm makes a payment that was promised in the contract, the lender does not bother with auditing anything. If the firm claims to have less than this promised amount, the lender audits (although in equilibrium it always finds that the firm was telling the truth!) and takes whatever is available.

Because this intuition is so simple and strong, the costly state verification (CSV) framework is tremendously widely cited in finance (and perhaps even more in accounting, since it is explicitly a model of auditing). As I mentioned earlier, the typical reference that people give is Townsend (1979), which is the most famous paper on this topic, even though his original paper is in some ways not the best model for a corporate finance setting.

CSV and the demand for safe assets

The CSV framework has a deep connection with a very important literature on why financial markets exhibit frequently manufacture "safe" assets out of risky ones. You could think of this as describing banking, corporate borrowing, securitization, and many other activities. It seems that the market's demand for any given investment is strictly less than the market's total demand for a safe tranche and a risky tranche of that investment, sold separately. Why?

One traditional argument is that the *risky* tranche is attractive to investors, as it gives them a way to lever up their returns. But it is increasingly easier for investors to get leverage on their own, yet this activity persists.

Instead, an influential literature argues that the *safe* tranche is particularly attractive, because it becomes a money-like investment. The CSV framework explains that, most of the time, you have very little incentive to investigate the quality of the issuer or collateral behind a low-risk asset. Importantly, everyone *knows* that you have little incentive. Then it becomes easy to transact trades

in this asset without concerns about adverse selection, and the asset starts to facilitate transactions, like money, even though it is not completely risk-free.

On the other hand, when any bad news arrives to sharpen everyone's focus on the quality of these "money-like" assets, the consequences can be disastrous as all trade freezes while everyone reassess their quality.

This model clearly describes the important traditional role of bank deposits in the economy. This literature has argued that basically the same economics can explain a much wider range of "safe" assets – both why they are subject to rare, sudden collapses in value with major spillover effects throughout the financial system, and why it would not be easy to get rid of this problem.

In my opinion this is one of the most important areas of active research, as it connects the long history of research on private money creation with the present and future of banking regulation and the payments system. A particularly clear and well-known reference is Gorton and Pennacchi (1990). That paper does not explicitly draw the connection with the CSV literature, but some of the later papers in this area do. In a future iteration of the course I will put together a much bigger section on this topic.

5.4 Conclusion

The basic question for this chapter was, why are debt contracts so pervasive in practice? We have seen two models of security design that present arguments to answer this question. In Innes (1990), debt is a second-best alternative to the live-or-die contract, which gives the entrepreneur the strongest possible incentives. In the CSV framework, debt is a way to avoid spending excessive resources on auditing a borrower's financial condition.

One obvious question is, if these models are to be believed, how could a firm ever issue equity? Or to put it differently, if the firm already has public equity outstanding, then the problems highlighted by these papers have been overcome in the past, so why should it be difficult to do the same again? Hence these models should generally be understood as describing very small firms, or in some abstract way as capturing forces that are important at the margin in a richer model, and may or may not be the dominant issue at any point in time.

Or, in the case of the CSV framework, we could argue that assessing the borrower's ability to pay its debt is a very different task from assessing the ability to pay dividends. Debt is a large payment due right now, dividends are a stream of payments that potentially last forever and can be postponed at little cost. You could perhaps imagine that a firm can persuade investors to trust its future long-term dividend payments, without necessarily persuading them to trust its near-term liquidity. Then there could still be a role for this model to generate the optimality of a debt contracts even for a large firm with publicly-traded equity.

5.5 Exam question: Optimality of debt with effort choice

Consider the following setup based on Innes (1990):

- An entrepreneur must raise financing of I to invest in her project.
- The project generates a random cash flow R that is distributed on $[0, \bar{R}]$ with density function $p(R|e)$, where e is the entrepreneur's choice of effort.
- To raise financing, the entrepreneur offers a contract to the investor, consisting of a payoff $w(R)$ for the entrepreneur and $R - w(R)$ for the investor. We impose a feasibility constraint that $0 \leq w(R) \leq R$.
- We also assume that the entrepreneur cannot commit in advance to any choice of e, but rather will choose it rationally ex post, given the contract.
- The lender's outside option is zero, so he will fund the project provided that the contract offers him at least I in expected value.
- The entrepreneur's objective function $\mathcal{U}$ is given by the expected payoff of the contract, minus a disutility of effort $g(e)$:

$$\mathcal{U} \equiv \int_0^{\overline{R}} w(R)p(R|e)dR - g(e)$$

  Assume that $g' > 0$, $g'' > 0$, $g(0) = g'(0) = 0$, $\lim_{x\to\infty} g'(x) = \infty$.

- Assume that p satisfies the *monotone likelihood ratio property* (MLRP):

$$\frac{\partial}{\partial R}\frac{\partial p(R|e)/\partial e}{p(R|e)} > 0$$

Given the above setup, the entrepreneur's problem is

$$\max_{w(\cdot),e} \int_0^{\overline{R}} w(R)p(R|e)dR - g(e) \tag{5.9}$$

subject to the constraints

$$\int_0^{\overline{R}} w(R)\frac{\partial p(R|e)}{\partial e}dR = g'(e) \tag{5.10}$$

$$\int_0^{\overline{R}} [R - w(R)]p(R|e)dR \geq I \tag{5.11}$$

$$0 \leq w(R) \leq R \tag{5.12}$$

Do the following:

1. Explain why the constraint (5.10) appears, and why it looks the way it does.

2. Following the steps below, show that the entrepreneur's optimal choice of contract takes the following "live-or-die" form: The entire project payoff goes to the entrepreneur, as long as it is above some endogenous threshold R^*; otherwise, the entire project payoff goes to the lender.

   (a) Write out the entrepreneur's problem in Lagrangian form, ignoring constraint (5.12), and attaching multipliers to (5.10) and (5.11).

   (b) Group all the terms involving w inside one integral. From this, explain why the optimal contract assigns the entire project payoff to one party or the other, depending on R.

   (c) Explain further why the optimal assignment rule is just a cutoff R^*, and why it is the *entrepreneur* who gets the project payoff when this payoff is high, and the lender who gets the payoff when it is low. Give both the mathematical justification, and some economic intuition, for why the outcome of the problem looks this way.

3. What additional constraint did Innes impose to conclude that a standard debt contract is optimal? Why did he argue that this constraint was reasonable?

4. How would the solution change if the entrepreneur was risk-averse? You can just describe this intuitively, without formal analysis.

Solution to exam question 5.5

1. This constraint captures the entrepreneur's behavior to choose effort level optimally ex post, given the contract that is specified. The left side is the expected marginal benefit of effort. The right side is the marginal cost of effort. If the equality is satisfied then the net marginal benefit of effort is zero. The other conditions on the problem guarantee that this first-order condition leads us to an optimum effort choice.

2. (a)

$$\begin{aligned}\mathscr{L} \equiv & \int_0^{\overline{R}} w(R)p(R|e)dR - g(e) \\ & + \mu \left[\int_0^{\overline{R}} w(R) \frac{\partial p(R|e)}{\partial e} dR - g'(e) \right] + \lambda \left[\int_0^{\overline{R}} [R - w(R)]p(R|e)dR - I - A \right]\end{aligned}$$

   (b)

$$\begin{aligned}\mathscr{L} = & \int_0^{\overline{R}} w(R) \left[1 + \mu \times \left(\frac{\partial p(R|e)/\partial e}{p(R|e)} \right) - \lambda \right] p(R|e)dR \\ & + \lambda \int_0^{\overline{R}} Rp(R|e)dR - I - A - g(e) - \mu g'(e)\end{aligned}$$

   The second line above does not depend on $w(\cdot)$. In the first line, the simple structure of the integral implies that we want to set w as high (low) as possible whenever the bracketed term is positive (negative). The lowest and highest values of w that we can set are 0 or R, so we simply assign the entire project to one party or the other.

   Within the bracketed term in the first line, μ and λ are constants, only one term depends on R. And our assumption of MLRP in the model setup was exactly that this term is monotonic in R. This explains why the assignment rule is just based on a cutoff value of R.

   (c) Our MLRP assumption was that the key bracketed term is not only monotonic but *increasing* in R. Hence, high values of R are where it makes the most sense to assign the project value to the entrepreneur. Economically, we assumed via MLRP that high cash flows suggest high effort levels, which makes sense. Then the optimal contract induces effort by promising high payoffs when effort was likely high.

3. Innes assumes the payment to the lender must be monotonic in R, based on a loose argument that a non-monotonic contract would create excessive incentive to hide and misreport cash flows.

4. The solution above relies very strongly on risk neutrality. If the entrepreneur was risk-averse, we would have to offer him a flatter wage contract as a function of R, and this would decrease his optimal effort level, illustrating the standard tension between "incentives and insurance."

5.6 Exam question: Optimality of debt with misreporting

Consider the following model, based on Gale and Hellwig (1985):

- A risk-neutral entrepreneur must raise financing of I to fund his project.
- The project's random payoff $R \geq 0$ is distributed with density $p(\cdot)$.
- When R is realized, the entrepreneur observes it, and can send a report $\hat{R}$ about it to the investor. The investor can also pay a cost $K > 0$ to directly observe R, but otherwise the investor will not directly observe R.
- Investment contracts consist of the following three functions of the report:
  - a decision whether to audit $y(\hat{R}) \in \{0, 1\}$,
  - a payoff function $L_0(\hat{R})$ to the lender when $y = 0$,
  - a payoff function $L_1(\hat{R}, R)$ to the lender in the case $y = 1$.

  The entrepreneur's payoff $w(\hat{R}, R)$ is project value net of the lender's payoff,

$$w(\hat{R}, R) \equiv R - y(\hat{R}) \times L_1(\hat{R}, R) - (1 - y(\hat{R})) \times L_0(\hat{R})$$

The entrepreneur chooses a contract to maximize this expected payoff, subject to the constraint that investors expect a return of I, net of audit costs.

We can apply the *revelation principle* (Myerson, 1985), and restrict attention to contracts where the entrepreneur tells the truth (sets $\hat{R} = R$ for all R), while attaching a constraint that this behavior is optimal.

Finally, we impose the feasibility constraint $0 \leq w(R, R) \leq R$. That is, L_0 and L_1 cannot specify payments that are negative, nor greater than R.

Then the entrepreneur solves

$$\max_{y(\cdot), L_1(\cdot), L_0(\cdot)} \int w(R, R)\, p(R)\, dR \tag{5.13}$$

subject to the constraints

$$\int [R - w(R, R) - y(R)K] p(R) dR \quad = \quad I \tag{5.14}$$

$$w(R, R) = \max_{\hat{R}} w(\hat{R}, R) \quad \forall R \tag{5.15}$$

$$0 \leq w(R, R) \leq R \quad \forall R \tag{5.16}$$

Assume without proof that a solution to this problem exists, and that this optimal contract specifies $y = 0$ on some interval of reports R.

Use the following steps to show that the optimal contract is a debt contract:

1. Prove that the optimal contract satisfies the following properties, in each case by contradiction: Suppose the optimal contract does not satisfy the given property, and explain why this would violate one of the constraints.

(a) $L_0(R) = F$ for some constant $F > 0$. That is, L_0 is a constant.

(b) For $R < F$, we have $y(R) = 1$.

(c) If $R > F$ and $y(R) = 1$, then $L_1(R, R) \leq F$.

2. Show that the original problem is equivalent to choosing y, L_1, L_0 to minimize the probability of audit, subject to the original constraints.

3. With the results from #1, and the reformulation from #2, prove by contradiction that the optimal contract must satisfy the following properties (except possibly on a set that has probability zero).

   (a) For $R > F$, we have $y(R) = 0$.

   (b) For $R < F$, we have $L_1(R, R) = R$.

   A process you can follow for #3 is:

   - For each statement, suppose you have a candidate contract that fails to satisfy the statement (on some set with positive measure), but satisfies all constraints and all the properties derived earlier.
   - Show that if you alter the contract to satisfy the statement, you generate a greater lender payoff, and no greater audit probability.
   - Then explain why it is possible to adjust F to a point where the lender's expected payoff is again I, but the audit probability is lower than in the original candidate contract.
   - Hence the original contract was not optimal.

   The mathematical details of this may become tedious. You do not have to spell out every detail formally, as long as your reasoning is clear.

We can then conclude that the optimal contract is a standard debt contract.

Solution to exam question 5.6

1. (a) Suppose not. Then there are R_1 and R_2 for which $y(R_1) = y(R_2) = 0$, such that (wlog) $L_0(R_1) > L_0(R_2)$. Then $w(R_2, R_1) = R_1 - L_0(R_2)$ is greater than $w(R_1, R_1) = R_1 - L_0(R_1)$, giving the entrepreneur the incentive to misreport and hence violating constraint (5.15).

   (b) Suppose $y(R) = 0$ for some $R < F$. From the previous point we have $y(R) = F$. But then $w(R, R) = R - F < 0$, violating (5.16).

   (c) Suppose not, then there is some $R > F$ for which $y(R) = 1$ and $L_1(R, R) > F$. Let R^* denote any report for which $y(R^*) = 0$. (We have not actually proved that there is any such report, but the question says we can just assume it.) Then $w(R^*, R) = R - F$ is greater than $w(R, R) = R - L_1(R, R)$, giving the entrepreneur to misreport and hence volating (5.15).

2. Restate (5.14) as

$$\int w(R, R)p(R)dR = \int [R - y(R)K]p(R)dR - I$$

   and substitute this directly into (5.13). We can ignore the constants I and $\int Rp(R)dR$ so the objective is just to maximize $-\int Ky(R)p(R)dR$, i.e. minimize $K \int y(R)p(R)dR$. We can then discard the constant K as well, and observe that the remaining integral is just the probability of audit.

3. (a) Suppose there is a region $Y_1 \subset [F, \infty)$ with positive probability measure, on which the optimal contract sets $y = 1$. We observe from the prior results that $L_1(R, R) \leq F$ on this region. Then we can strictly increase lender expected payoff, and strictly decrease audit probability, with a contract that sets $y = 0$ for all $R > F$ (and $L_0 = F$).

   Finally, observe that the lender payoff from this modified contract is a continuous function of F, and it equals zero when $F = 0$. By the intermediate value theorem we can decrease F to some lower value that will generate a payoff of exactly I to the lender, and will have strictly lower audit probability than the contract that we started with. Then the original contract was not optimal.

   (b) Suppose there is a region $Y \subset [0, F]$ with positive probability measure, on which the optimal contract sets $L_1(R, R) < R$. Then we can strictly increase lender payoff by setting $L_1 = R$ on this region, without affecting audit probability.

   Finally, observe that the lender payoff from this modified contract is a continuous function of F, and it equals zero when $F = 0$. By the intermediate value theorem we can decrease F to some lower value that will generate a payoff of exactly I to the lender, and will have strictly lower audit probability than the contract that we started with. Then the original contract was not optimal.

Chapter 6

Banking

6.1 Overview

Banks are an obvious topic of interest to financial economists. After all, finance is the study of how funds are connected to investment opportunities, and banks historically are a critical link in this chain. Banks are also some of the largest companies in the portfolio of the average stock market investor, and are some of the primary employers of the students who take our finance classes.

However, it is not immediately obvious why banks should need their own class of *models*. At first glance, they are businesses like any other, engaging in an activity (financial intermediation) to generate profits for their shareholders (or partners, or other claimants, depending on how they are organized). If anything, banks seem like some of the most ruthlessly profit-maximizing organizations in the economy, and might seem to fit standard models unusually well. Hence, the real point of banking theory is to study the ways in which banks *are* quite different from any other business.

The most fundamentally unique feature of banks, compared to any other business, is their funding structure. Throughout history and in every society, banks' main source of funding is *deposits*, which are then used to make risky loans (to businesses or to households) with the expectation of profits. Hence the bank's balance sheet mainly consists of deposits as a *liability*, and its loans or other investments as an *asset*.

The details of deposit contracts vary from one setting to another, but one feature is universal: Banks promise depositors the ability to retrieve their funds from the bank, more easily than the bank itself can retrieve the funds from its borrowers. Obviously, this fact is appealing if everyone expects the bank to honor its promises, but can lead to great danger if not!

For example, in the United States, a major traditional activity of banks is to use customers' *demand* deposits, which can be withdrawn at any time whatsoever, to originate long-term loans that fund local businesses or home purchases. The bank will also keep a large amount of cash on hand to meet depositor

withdrawals day-to-day. But, if *all* depositors hypothetically demanded their deposits back at the same time, that cash would quickly be depleted.

What can the bank do in response? It cannot demand cash from its loan customers. It can try to sell those loans for cash to some other investor, but most likely the sale price will be less than the total amount of the bank's deposit liabilities. The bank can try to suspend the payment of deposits (by closing for business, shutting down its website, etc), but this violates the deposit agreement and cannot last long. In the end, the bank might have to declare bankruptcy, leaving some depositors with losses, if it cannot find an outside rescue.

This situation is known as a *bank run.* Bank runs and bank failures recur throughout history, and are a powerful part of the cultural memory in many societies. For many people, their bank deposits are one of the only interactions they have with the financial system, and they tend to imagine (understandably) that those deposits are equivalent to cash. The sudden failure of a bank without any warning is a deeply disturbing and traumatizing event, even in the modern world where most customers' deposits are insured, and much more so in past times where they might have suffered severe losses.[1]

An especially striking feature of bank runs is that customers often respond more to their fears about *other customers' behavior* than to their own concerns about the bank's health. The bank may be fundamentally "solvent," in the very specific sense that it will have plenty of cash at every date (now and in the future) to meet depositors' actual *need* for cash on those dates. Then there would be no problem if all depositors'would simply agree not to withdraw their *unneeded* deposits today. And depositors themselves may understand this fact perfectly well! Nevertheless, the bank run can happen anyway, simply because depositors think that *others* plan to withdraw first, leaving them with nothing. Hence bank runs depend critically on depositors' psychology and beliefs, which may explain why they often seem to develop spontaneously without warning.[2]

One fascinating fact in the modern financial system is that the above description applies far beyond the classic setting of a deposit-taking bank. Over and over, financial intermediaries arise that promise liquidity to customers while investing their funds in fundamentally *illiquid* places. The most prominent examples (but far from the only ones) are open-ended funds, such as prime money market funds, bond mutual funds, and many REITs. These funds offer investors the ability to redeem at a nominal value that is not necessarily the same as what the fund itself could raise by selling its assets. The predictable result is that when the underlying assets lose value compared to that nominal promise, cus-

[1] In the US, the films *It's a Wonderful Life* and *Mary Poppins* include depictions of bank runs that are especially well-known, and are often referenced in popular discussions of banking.

[2] This was vividly illustrated by the sudden collapse of Silicon Valley Bank and other regional banks in 2023. The banks' assets were mainly Treasuries, MBS, and other investments with extremely low credit risk. It was quite clear that if held to maturity, these assets would generate plenty of cash flow to meet depositors' future needs. But the rapid rise of interest rates meant that the liquidation value of those assets would not be enough to meet requests *if everyone decided to withdraw immediately.* This was widely understood for several months and did not appear to be a problem – then customers suddenly changed their minds, and the banks collapsed quickly.

tomers may suddenly decide to cash out as quickly as possible at the inflated valuation, out of fear that others will do so first. The fund suspends redemptions if it is able, otherwise it quickly collapses.

The bank-run dynamic is not only terrifying for the customer, but also seems to have severe consequences for society at large. Economies often recover surprisingly quickly from dramatic shocks like war, pandemic, natural disaster, and so on. On the other hand, the mere presence of *doubts* about the solvency of institutions in the financial sector (banks, mutual funds, etc) often lead to severe and protracted slumps in economic activity, that seem vastly disproportionate to the amount of losses actually *experienced* by those institutions.

With all these points in mind, the key questions of banking theory become clear. What exactly causes a bank run? Why do concerns about bank health seem to lead to such destabilizing effects on the macroeconomy? Why do financial intermediaries keep setting up structures where such things can happen, as opposed to stabler alternatives like closed-end funds or the typical equity-financed corporation? Are such structures good for society since they are able to fund large amounts of investment most of the time, or bad since they require expensive bailouts and deposit insurance schemes? Should we be particularly concerned about the moral hazard problems created by those policies?

Despite all this motivation, we only have time to study the very tip of the iceberg. We will focus on two of the most widely-studied models of banking:

- Diamond (1984) provides an argument for why banks offer deposit contracts, despite the risk that they entail. The key observation is that the deposit contract is a form of debt, which provides maximal incentives for the bank to report investment outcomes truthfully, following the arguments from the literature on costly state verification (see Chapter 5). Thus, fragile deposit contracts are in fact a deliberate feature of the banking system.

- Diamond and Dybvig (1983) provides a formal description of the bank run phenomenon, based on a self-fulfilling investor belief that is independent of the banks' actual health. The main point is that a bank financed by demand deposits inherently features multiple equilibria: One in which the outcome is efficient, and one in which the bank immediately fails.

These models are two of the major workhorses in the banking literature. On their own, they do not give much policy advice, but their underlying descriptions of the banking sector are often embedded in richer frameworks that do, and their arguments have also been extended in many other directions.

I will mention a few things to keep in mind as you read:

Most people take it for granted that fractional-reserve banking is a critical feature of a developed economy, that the instability of such banks is fundamental and unavoidable, and that we can only hope to manage this through policy interventions, while accepting the inevitable drawbacks of those policies. However, that has never been universally accepted. After the Great Depression, the so-called "Chicago Plan," endorsed by many prominent economists, called for a

full-reserve banking system. The basic idea is that any deposits (i.e. promises to repay on demand or on short notice) would have to be backed 100% by truly safe assets (such as reserves with the central bank). Any investment in assets with credit risk or interest rate risk would have to be accomplished through vehicles whose liabilities are equity or long-term debt, not deposits. The idea persists today in the calls for "narrow banking," and it has always had the support of at least some high-profile economists. Those who oppose it generally argue that such a system would feature much less volume of lending and investment than what we currently have. Since it has never been tried, the only way to analyze such arguments is with a clear model that explains why banks look they way they do in the first place, and whether the externalities of their business model are a net positive or negative for the rest of the economy.

A related perspective on these issues connects with our capital structure topic from earlier chapters. Deposit liabilities should be understood as a form of debt, and from this perspective, banks are by far the most leveraged businesses in the economy, with equity value that is commonly around 10% or less of total assets (compared to well over 50% for the typical nonfinancial corporation). In general, any organization that funds long-term illiquid assets with short-term liabilities having a fixed nominal value is subject to problems like those described here. So to connect with our earlier topics, one could view the bank's problem as a particularly severe case of debt overhang and asset substitution problems. Then, a common academic perspective is that the cleanest way to regulate banks is simply to require them to issue far more equity than they currently do.[3]

One of the clearest arguments for why the traditional bank structure may in fact be socially *desirable* comes from the literature, mentioned at the end of the prior chapter, on the demand for "safe" assets. The idea is that economic agents need a certain amount of safe assets to be used in negotiating transactions, and bank deposits can fill this role (though other assets can too). Under this view, if we banned deposit-taking banks from making risky investments, consumers will still keep roughly similar amount of deposits with banks as they currently do, but we would be sacrificing investment on a large scale. Whether this is true, and whether it would be less efficient than our current world, is again a question for theoretical analysis. This line of thinking connects banking theory with the perspective that banks "create" money by originating loans with matching deposit liabilities, and hence to even deeper questions about the nature of payments systems in the economy, the optimal quantity of money, and whether the government should monopolize money creation, or allow it to be yet another activity accomplished within a competitive private sector.

6.2 Diamond (1984)

Diamond (1984) gives an argument why the structure of a bank may be a

[3]Bank representatives tend to vehemently oppose such ideas. There may be valid reasons for this, but the arguments they present are often obvious fallacies based on the logic of Modigliani and Miller (1958). Anat Admati has been especially vocal about this issue.

desirable, even optimal way to finance investment, despite the fragility created by its run-prone nature. Indeed, this paper will argue that fragility actually plays an important and valuable role in maximizing the value created by the banking sector.

The basic idea is built on the models of costly state verification, as described in Chapter 5. The argument is that if the optimal lending contract imposes costs of state verification in the event of default, then it would be inefficient for many individuals or small entities to provide these contracts and bear these costs. Instead, there is value in pooling all investments into a large entity, which we call the bank, that does all the investing. This system can economize on the amount of monitoring costs that would otherwise be borne in aggregate. The result is a "delegated monitoring" theory of the banking system.

The argument is a bit more subtle than it appears at first glance. The above reasoning does not quite explain why the bank is better from a social perspective, since it sounds like the same amount of monitoring is happening overall. A critical part of the reasoning will be to assume that each individual depositor is small relative to the size of each individual project to be financed, and hence each project will raise capital from multiple investors. Without some form of coordination between them, each individual investor might have to bear the costs of state verification in the event of default, which would be redundant and wasteful. The role of the bank is to be a single representative for all investors who can pay just a single verification cost (per project) on all of their behalf.

This in turn raises the question of how the investors can trust the bank to repay their funds, if they could not trust the manager of the project to do so. The paper's answer is that investors will still employ a debt contract when they provide their funds to the bank. We interpret this debt contract as a deposit contract since both specify a fixed redemption value that is a senior liability. Investors now pay the cost of state verification when the bank defaults on their deposits, rather than when projects default on loans.

This might seem even worse than before, since the system now features monitoring costs in multiple places! But a key insight is that, by diversifying across a growing number of projects, the bank can drive its default probability down towards zero. This minimizes the costs that depositors incur on average, leading to an overall cheaper system.

6.2.1 Model setup

Agents Everyone is risk-neutral. There are a large number of entrepreneurs with no wealth, but with the ability to operate a project. Each project requires 1 unit of wealth and investment, and will generate a random cash flow $\tilde{y}$. The probability distribution of $\tilde{y}$ is common knowledge and does not depend on any action by the entrepreneur. There are also a large number of investors each with $1/m$ units of wealth where $m > 1$. Hence a given project can be funded only if it attracts m investors. The investor's opportunity cost is to earn a rate of return R on each dollar of their funds.

Information and contracting environment As in the "costly state verification" literature from the prior chapter, the only person who can always observe a project's cash flow is the entrepreneur of that project. This creates a problem for the outside investor of how to enforce repayment.

In particular, we assume that each project has a strictly positive probability of $\tilde{y} = 0$. So if the entrepreneur reports that there is no cash flow to support repayment, the outside investor will not necessarily know that the entrepreneur is lying. As a result, there is no way to enforce a contract that specifies *any* positive cash flow from the entrepreneur to the investor.

However, while agents cannot enforce contracts that depend on the realization of $\tilde{y}$, they *can* enforce contracts that depend on the cash flow z that the entrepreneur offers to the investor. In particular they can write a contract that imposes a disutility $\phi(z)$ on the entrepreneur depending on the cash flow he offers. This disutility ϕ is *not* a transfer to the investor, but rather just some pain that the entrepreneur will suffer.

We can immediately see that the loan parties will optimally set ϕ to be decreasing in z. The next section expands on this intuition and shows that debt is the optimal contract.

6.2.2 The optimal contract without monitoring

We define the optimal contract as the solution to the following problem:

$$\max_{\phi(\cdot)} \mathbb{E}[\tilde{y} - z - \phi(z)]$$

subject to

$$z \in \arg\max_{z \in [0,y]} y - z - \phi(z)$$
$$\mathbb{E}[\tilde{y} - z - \phi(z)] \geq R$$

In words, the entrepreneur's expected payoff from any contract is the expected project cash flow, net of repayment and the penalty ϕ. The entrepreneur offers a contract ϕ that he chooses to maximize this payoff, subject to the constraints that (1) he will choose the repayment z that seems optimal ex post, and cannot commit to do otherwise, and (2) the investor expects a return of at least R.

Proposition 6.1 (cf Prop 1 of Diamond, 1984)**.** *The optimal contract sets* $\phi(z) = \max(h - z, 0)$ *where* h *is the smallest value such that* $\mathbb{E}[\min(\tilde{y}, h)] = R$*.*

Proof. See Diamond (1984). □

The structure of this contract is intuitive but slightly disturbing: We determine a face value of debt h such that the investor breaks even in expectation, and then impose nonpecuniary penalties such that the entrepeneur *will* pay h no matter what, either in money or in disutility. This of course requires that it is actually possible to impose the punishment h. In practice there are limits on the punishments that investors are able or willing to impose. This would constrain the settings in which this model's optimal contract is feasible.

6.2.3 Delegated monitoring

Section 3 of the paper introduces the delegated monitoring problem:

The intermediary is risk-neutral and has no wealth. It raises funds from the original lenders (whom we now call depositors), invests those funds in the projects, and can monitor those projects. Since depositors each have $1/m$ of capital and each project requires 1 unit, an intermediary who funds N entrepreneurs must have $m \times N$ depositors.

We assume that the depositor-intermediary relationship faces the same reporting problem as the investor-project relationship: After realizing payoffs from the projects, the intermediary reports the total payoff to the depositors, and depositors cannot take it for granted that the intermediary will tell the truth. We will assume that depositors solve this problem by using debt contracts with nonpecuniary penalties, following exactly the same logic as before.

Consider a population of N entrepreneurs funded by the intermediary. An intermediary who funds N entrepreneurs must promise $N \times R$ to depositors in expectation. Ex post, the intermediary will receive $G_N = \sum_i g_i(y_i)$ from entrepreneurs, where $g_i(y_i)$ denotes the payment made by the i-th entrepreneur, and will pay $Z_N \leq G_N$ to depositors. Both G_N and Z_N are realizations of random variables, with $G_N \in [0, \overline{G}]$.

As before, we conclude that the optimal contract applies nonpecuniary penalties, now labeled $\Phi(Z_N)$, in the event of default. By the same argument as in Proposition 6.1, the optimal contract sets a face value H_N that allows depositors to break even in expectation, and sets $\Phi(Z_N) = \max(H_N - Z_N, 0)$ to guarantee that the intermediary *will* pay depositors H_N no matter what.

Now we move to the date in the model *after* such a contract has been signed, so H_N is given, and we ask ourselves what kind of monitoring decisions the intermediary will make. At this point the intermediary faces a payoff of $\mathbb{E}[G_N] - H_N$, and is unable to change H_N, so its goal with its monitoring decision is simply to maximize $\mathbb{E}[G_N]$. This puts us back in the framework of the earlier section, and by that argument we conclude the intermediary will use a debt contract for each entrepreneur with the structure we derived as before.

Thus we have derived a debt-financed banking system: Investors (depositors) provide their funds to a single intermediary (the bank), with the contract taking the form of debt (which can be interpreted as a deposit contract since it specifies a fixed redemption value). The bank in turn funds all the projects.

6.2.4 Discussion

At first glance, it may not be clear if this system is better than having the depositors invest directly in the projects. On the one hand, any individual project that defaults will only result in one cost of verification (by the bank), whereas this would have resulted in $m \times N$ such costs before (since there were m investors per project). On the other hand, there is now the risk of monitoring costs between depositors and banks: If the bank turns out to be insolvent, *every one* of the $m \times N$ depositors will incur their own cost to verify the state. Based

on this tradeoff, at small scale (i.e. sufficiently small N) the banking system may be less efficient than direct investment.

But as N grows, the banking system becomes more efficient, due to a key *diversification* effect. Usually in finance, diversification is valuable as a way of sharing risk across individuals, but that is not the idea here since everyone is risk-neutral. Instead, diversification in this model is a way of economizing on monitoring costs, and specifically on the costs between the depositors and the bank. For the bank to be insolvent would require a critical mass of projects to default. As the total size of the banking system grows, the probability of this happening limits to zero, and we realize only the benefit of the banking system that was described in the previous paragraph, not the cost.

Proposition 6.2 (cf Prop. 2 of Diamond, 1984)**.** *Suppose entrepreneurs' projects returns follow identical, independent, bounded distributions. Then the total cost of delegation, per entrepreneur monitored, approaches zero as* $N \to \infty$.

Diamond (1984) has become very influential in discussions of the banking system, because it provides a compelling argument that nicely fits the major stylized facts about that system, and that generates a clear (though perhaps uncomfortable) policy interpretation. The message is that banks are *deliberately* fragile as a way of imposing discipline, and attempting to remove fragility might inadvertently remove that discipline as well.

The idea of fragility as a disciplining device has been extended in other ways. Another influential example is Diamond and Rajan (2001). In that paper, fragility again disciplines the bank, but the action that we are concerned about is different. In Diamond (1984), the goal was to induce the bank to report project cash flows truthfully. Here, we assume the bank has a special skill in helping to maximize the borrower's value, and the challenge is to encourage the bank to exert maximal effort at this task. To justify why the bank has such skill, the implicit assumption is that they learn about business conditions and market opportunities through day-to-day operations in their branches. This complementary between deposit-taking and loan-making has been another highly influential part of Diamond and Rajan (2001).

The overall message of these two models, and many others in this area, is that the fragility of the deposit contract provides banks with very strong incentives to maintain cash flow. From this perspective, perhaps the basic structure of a bank keeps arising in practice because it is an easy (if imperfect) mechanism to overcome a specific type of hidden-action problem.

Of course, even if we accept this narrative, we do not have to be happy about it! Most depositors would probably be uncomfortable with the idea that their contracts are designed to make banks fragile and therefore cautious. But having identified this valuable role of the deposit contract, perhaps we can guide future research toward finding better practices to achieve the same goal.

6.3 Diamond and Dybvig (1983)

This is without question the best-known paper on banking theory. The main goal of the paper is to *describe* what bank runs are and why they happen. Unlike Diamond (1984), it does not try to argue that any particular part of the banking setup is optimal. The paper is extremely clear and worth reading in full, and we will follow its exposition closely.

The paper's biggest contribution is to view bank fragility as a problem of *multiple equilibria*. In the model, there will be one equilibrium in which the bank operates smoothly, taking deposits and funding risky investments; and another in which the bank collapses. Because the two equilibria coexist, it seems somewhat arbitrary which of them will come to pass. The bank run is interpreted as a sudden shift between the two, happening for some arbitrary reason outside the model. In this sense, multiple equilibria become a prediction of the model rather than a weakness of it.

The most provocative point is that the bank is equally healthy in both equilibria, so the run is a pure "panic" that arises only due to self-fulfilling psychology and beliefs by consumers. This aspect of the model has been very divisive. Many researchers find it to be plausible and realistic. Others argue that real-world bank runs are always connected to some "fundamental" news about the bank's health, and point out that the contrary result in Diamond and Dybvig relies on some questionable choices about how to model banking itself.

This controversy between the "panic" versus "fundamental" view of bank runs has continued to the present day. Some of the most influential research in banking theory has shown how to synthesize the two together. In this approach, runs are triggered when investors receive negative news about the bank's health but then overreact to it. This approach requires technical tools from the "global games" literature, so unfortunately we will not be able to do it justice in our course, but I will try to summarize the intuition at the end of this section.

6.3.1 Model

There is a continuum of consumers who start the model with an endowment of 1, and will have no further endowment beyond that date. Consumers are identical at $t = 0$, but will find out at $t = 1$ that they are one of two types:

- W.p. π, they are type 1 ("impatient") and consume only at $t = 1$.
- W.p. $1 - \pi$, they are type 2 ("patient") and consume only at at $t = 2$.

Consumers have log utility at the date when they consume, and apply a discount rate of $\rho < 1$ between dates. Consumers can store cash for zero return, the main implication being that a type-2 consumer who receives cash at $t = 1$ will simply hold it until $t = 2$, then consume it.

Then, if we let y_1 and y_2 denote the *income* of an agent at dates 1 and 2, the type 1 consumer gets utility of $\ln(y_1)$, and the type 2 consumer gets utility of $\rho \ln(y_1 + y_2)$. At $t = 1$, agents will act to maximize their known utility function.

At $t = 0$, they will act to maximize their expected utility, attaching probability π to being type 1 and $1 - \pi$ to being type 2.

Finally, we describe the investment opportunity that all consumers have:

- At $t = 0$, an investment of 1 can be made in a long-term project.
- If left until $t = 2$, the project will yield a payoff of $R > 1$.
- The project can also be liquidated at the earlier date $t = 1$, but then it will only pay back the initial investment of 1.

6.3.2 Results

Without the bank

Before introducing the bank, we ask what agents will do if left on their own and unable to transact with each other (the standard term for this is "autarky"). This is simple: Everyone will invest in the project initially, but, those who draw type 1 will liquidate their project early and get consumption of 1. Everyone's expected utility is $(1 - \pi)\rho \ln(R)$.

For comparison, we can also determine the *optimal* outcome in this model. Agents are exposed to a large amount of risk (the uncertainty in whether they will be type 1 or 2). This matters greatly to them as the two agent types have a vast disagreement in their desired outcome. But since everyone has symmetric exposure to these risks, and we know the fraction of agents who will ultimately be each type, it seems reasonable that they could achieve a better outcome if they shared risks.

To see what this scheme would look like, we can set up an explicit maximization problem for a benevolent social planner who can allocate any available consumption ex-post according to type:

In principle there are four levels of consumption to choose: the consumption for types 1 and 2 at date 1, c_{11} and c_{21}, and for types 1 and 2 at date 2, c_{12} and c_{22}. However we can immediately observe that the *optimal* rule will set $c_{12} = c_{21} = 0$ because type 1 gets no utility from consuming at date 2, while type 2 would always prefer to store consumption for another period at any positive rate of return rather than receive it at $t = 1$. Hence we will just relabel c_{11} and c_{22} as c_1 and c_2, and consider the optimal choice of these values.

There is a resource constraint on the problem: There is a unit continuum of consumers at $t = 0$, each with 1 unit of endowment, and these endowments must finance all investment. There will be measure π of agents who receive c_1, and $1 - \pi$ who receive c_2, so we will need to make sure that we have available at πc_1 by date 1, and that our long-term investments generate $(1 - \pi)c_2$ by date 2. The first of these requires us to allocate πc_1 units of endowment already at date 0, since we cannot earn any positive return before date 1. The second only requires us to allocate $(1 - \pi)\frac{c_2}{R}$ at date 0, since we will earn $R > 1$ by date 2. Altogether the resource constraint is

$$1 = \pi c_1 + (1 - \pi)\frac{c_2}{R}$$

The planner's problem is to maximize each consumer's expected utility:

$$\max_{c_1,c_2} \pi \ln c_1 + (1-\pi)\rho \ln c_2$$

Solve the constraint $c_2 = \frac{1-\pi c_1}{1-\pi} R$, substitute in and find the solution:

$$c_1 = \frac{1}{\pi + (1-\pi)\rho} \text{ and } c_2 = \frac{\rho R}{\pi + (1-\pi)\rho}$$

The important things to note here are $c_1 > 1$, $c_2 < R$, and expected utility is greater than it was under autarky. In other words, consumers all benefit from some degree of risk-sharing.

Nevertheless, as discussed on p.406 of the paper, agents cannot achieve this outcome by transacting directly with each other, if the only contracts available are risk-free bonds (i.e. one contract that pays out 1 at $t = 1$, and another contract that pays out 1 at $t = 2$). In fact, such contracts will not allow any improvement at all over autarky. At $t = 0$ there is no reason for anyone to trade with each other, since they don't yet know their types, and at $t = 1$ it is too late since the type 2 agents no longer have any incentives to trade at all. In a sense this conclusion is obvious: Risk-sharing must happen before the risks are realized, and, it requires contracts that are contingent on those realizations.

The bank enters the model as a type of insurance scheme that will move us closer to the optimal outcome, *without* introducing any contracts that depend directly on anyone's type.

The bank

Now we introduce a bank. Consumers can deposit their unit endowment with the bank at $t = 0$, and (by assumption) the bank offers them the following contract: At $t = 1$ they can withdraw an amount r_1 if they choose. Then at $t = 2$, the bank's remaining assets will be divided equally among anyone who has not yet withdrawn.

We want to allow for $r_1 > 1$, which raises the possibility there are more withdrawals at $t = 1$ than the bank can actually honor. In this situation, it is assumed that a random subset of withdrawal requests will be honored in full, while the rest are left with nothing. This is meant to capture what would happen in a richer dynamic model, where the bank keeps its promises as long as it can, hoping to avoid failure, and does not know until it is too late that it is indeed receiving more requests than it can honor. This critical aspect of the model is often called a *sequential service* assumption, reflecting the motivation.

The result of these assumptions is a *strategic complementarity* in depositors' payoffs. Let f denote the measure (fraction) of agents who choose to withdraw at $t = 1$, and consider the payoff to an individual agent depending on her own choice to withdraw or not, as a function of both f and the deposit contract r_1.

If she chooses *not* to withdraw at $t = 1$, her payoff will be

$$V_2(f, r_1) = \max\left\{0, \frac{1 - r_1 f}{1 - f} \times R\right\}$$

In words: If the bank did not run out of funds at $t = 1$, then, the total amount of resources left at $t = 1$ after paying out r_1 to all redeeming depositors was $1 - r_1 f$. This grows at rate R, and then is divided up at $t = 2$ among the $1 - f$ depositors who did not withdraw. On the other hand, if $1 - r_1 f < 0$ then the bank ran out of funds at $t = 1$ and this individual depositor is left with nothing.

Now consider if this depositor *does* withdraw at $t - 1$. If the bank is able to meet all withdrawal requests, $r_1 f < 1$, then this depositor simply gets r_1. But if $r_1 f > 1$ then the bank will fail at $t = 1$. Due to the sequential-service assumption, the depositor's payoff in this scenario depends on her place in line. More precisely, her probability of being early enough in line to withdraw her 1 unit is $1/r_1$, so her expected payoff to withdrawing in this scenario is 1. Altogether, the expected value of withdrawing at $t = 1$ is

$$V_1(f, r_1) = \begin{cases} r_1 \text{ if } f < \frac{1}{r_1}; \\ 1 \text{ if } f > \frac{1}{r_1} \end{cases}$$

Equilibria

Whenever there is strategic complementarity (whenever agents' payoffs depend directly on other agents' actions), then, we should often expect to find multiple equilibria. Indeed that is the main conclusion of the paper.

Consider if we try to achieve the first-best outcome using the bank setup. That is, set r_1 equal to our optimal value of c_1 from earlier.

The "good" equilibrium is one in which the only agents who withdraw are those who draw type 1. Then $f = \pi$. Since our optimal value of c_1 was less than this, we have $r_1 f < 1$ and the bank is able to meet all withdrawals at $t = 1$. Then the payoff to the type-2 agents is $V_2 = \frac{1 - r_1 \pi}{1 - \pi} \times R$ which exactly coincides with our optimal solution from earlier. Are all agents happy with their specified behavior? The type 1 agents clearly are happy withdrawing at $t = 1$ since they get no utility from consuming after this date. The type 2 agents, if they withdrew at $t = 1$, would simply store the funds r_1 for one date and then consume them, hence getting strictly less than if they waited to withdraw. So they also prefer to follow their specified behavior and leave their funds in the bank until $t = 2$.

We conclude that the first-best outcome is a Nash equilibrium of this model. This is remarkable! The main obstacle to achieving the first-best outcome through competitive markets was that no one could sign contracts contingent on their type (i.e. could not trade Arrow-Debreu claims). Yet the bank is able to overcome that obstacle despite not having any more information or technology than anyone else in the model.

But this conclusion comes with a dark side. Suppose we specify an alternative equilibrium, in which $f = 1$. This is indeed a Nash equilibrium: The type-2 agents get an expected payoff of only 1, but, *given that everyone else withdraws* at $t = 1$, their alternative to wait until $t = 2$ leaves them with nothing at all. Hence, if this is the equilibrium behavior that we specify, they will indeed cooprate. This is what we can describe as a bank run. The bad equilibrium

exists for *any* $r_1 > 1$, not just the "first-best" value.

Critically, there is no *fundamental* reason for the bank run to happen in this model (see for example pages 403-404). The good and bad equilibria happen with exactly the same model setup and parameter values: whenever the bad equilibrium happens, the good one is perfectly feasible as well. The bank run is triggered because depositors suddenly change their opinion, not about the *bank*, but rather about *each others' behavior.*

The bad equilibrium is particularly depressing, because it actually leaves everyone worse off in expectation than if they had simply remained in autarky! In autarky, you at least had the possibility of drawing type 2 and consuming the larger project return, such that expected utility was $(1-\pi)\rho \ln R > 0$. In the bad equilibrium, everyone gets a payoff of 1 for sure, so expected utility is zero. Hence, the paper does *not* show that the banking setup is clearly better than autarky, let alone better than alternative ways of financing the investment.

6.3.3 Discussion

Multiple equilibria as a model prediction

In the earlier chapters, we have implicitly regarded multiple equilibria as a weakness of our equilibrium concept. We have sought ways to strengthen that concept and pick one outcome as the "prediction." But here, the multiplicity of equilibria is in fact the central *prediction* of the model, and indeed it feels like a correct description of the bank-run scenario in reality.

To be clear, in principle there should still be some way to sort out which equilibrium will happen, and when exactly we will switch between them. (See discussion of global games below.) But such an analysis would still preserve the central observation here, that the exact outcome is at least somewhat arbitrary, relative to economic fundamentals. In this sense, the original model with its multiple equilibria feels perhaps *incomplete*, but not fundamentally *incorrect*.

Policy proposals: Deposit insurance and lender of last resort

Diamond and Dybvig (1983) consider two government policies that could prevent the bad equilibrium from arising:

The first is deposit insurance. The government forcibly taxes households in the economy, and it uses the funds to guarantee that anyone who waits until $t = 2$ will receive the full amount that they would have expected in the good equilibrium. This removes the incentives for type-2 consumers to liquidate early, regardless of what others are doing, and restores the good equilibrium.

The second is for the government to act as a lender of last resort. In some ways this is similar to deposit insurance: The government promises to lend to banks at a fixed interest rate, in any amount they might need. The bank will gladly borrow at this rate to meet any withdrawals at $t = 1$, as this allows its projects to carry through until date $t = 2$, providing surplus that can be used to repay the government loan and still realize a profit for the bank.

The striking thing about either of these schemes is not only that they eliminate the bad equilibrium, but also that (for precisely this reason) they *never actually have to be used.* Once the type-2 agents know that the bank will definitely be able to meet their withdrawal request, they have no reason to withdraw at all. In other words, deposit insurance is much cheaper than the nominal value that it promises to pay, because the promise itself guarantees that it will never *actually* need to be paid.

So, either of these schemes can make sure we end up only in the good equilibrium, which in turn is the first-best outcome, without actually spending any resources. It almost seems too good to be true! Have we solved the problem of banking regulation?

Not so fast. These schemes are a panacea *in this model*, because this is a model in which, by assumption, banks are always solvent in the long-run and the first-best is always feasible. The obvious concern with government-backed guarantees of any sort is that the government might end up guaranteeing *insolvent* banks. And these concerns are amplified if we think that banks themselves might begin to take into account the government guarantee, and begin to take excessive amounts of risk or to make low-quality loans – either deliberately, or simply by failing to screen and monitor borrowers as aggressively as they otherwise would. This is a form of moral hazard problem.

It is theoretically easy to understand the moral-hazard concerns with bank support programs of any kind. Researchers remain divided on how important such concerns are in practice. For our purposes, the important thing to realize is that such concerns do not appear in the model described in this chapter. Then it is not surprising that there is no apparent downside to schemes like deposit insurance or the lender-of-last-resort program.

The sequential-service assumption

The sequential-service assumption at $t = 1$ is an absolutely criticial ingredient of the main result. The strategic complementarity between investors only arises because the bank has made promises at $t = 1$ that it cannot actually honor if *everyone* asks it to. If it made a different promise (by using a different contract), the problem might not arise.

For example, the bank run can be avoided if the bank can simply ignore a high volume of withdrawal requests. Indeed, a traditional strategy for a bank to survive a run is to simply close for the day and stop answering the phones.[4] In the moment, this is a matter of survival. But if the bank could convince everyone *beforehand* that it would be able to shut off withdrawals during a run, then this would solve the problem. DD analyze this "suspension of convertibility" approach in their Section III: Formally, the bank modifies the deposit contract to allow only the fraction π of assets to be withdrawn at $t = 1$. Now there is no longer a bank-run equilibrium. We can see similar schemes in some investment vehicles like hedge funds that have rules limiting withdrawals at any moment,

[4] In modern bank runs like the one at SVB, websites and apps for handling electronic withdrawal requests often conveniently crash when receiving a high volume of requests.

but they seem infeasible in retail banking, as they require the bank to know exactly how many people "really" need their money at any point in time.

A similar, simpler way to avoid the bank run if the investment is financed in the first place through equity instead of deposits. That is, we separate the bank into a deposit-taking entity that maintains full reserve backing, and an investment-making entity that is financed with equity. Anyone who liquidates their investment in the latter entity at $t = 1$ is only promised their share of what is actually available. This removes any spillover effect or strategic complementarity between investors, and hence removes the bank-run equilibrium. This is essentially the motivation behind the "Chicago plan" and "narrow banking" approaches to the banking system.

"Panics" versus "fundamentals" and the role of the bank

We can describe the paper's big-picture story in a weaker and a stronger form. The weaker form is that *bank runs happen because deposit contracts lead to strategic complementarities between investors.* I think it's fair to say that this statement is now widely accepted, and in large part due to this paper.

The stronger form is that *bank runs happen for no fundamental reason.* That is, they are entirely a self-fulfilling "panic" by depositors, and nothing to do with the "fundamentals" of the bank itself. This is literally true in the model, but is much more controversial as a statement about reality. At least anecdotally, bank runs seem to be triggered by investor concerns about the bank's losses, contrary to the story of the paper.

To understand how the model generates runs based on pure depositor panic, we have to examine its underlying theory of how the bank works. This turns out to be controversial as well. To highlight just one issue (because I think it's the most important): The model's result seems to go away if banks offer the kind of payment services that are in fact one of their primary functions in practice.

In the model, when depositors need to finance early consumption, they must withdraw *real* resources from the banking system. The most literal interpretation of this is that the bank has financed a farm, and the depositor demands a harvest before the crop has matured, ultimately damaging the overall yield. A more realistic interpretation is that the depositor must pay for their consumption in physical currency, and so they demand that the bank redeem their deposits with this currency. Currency cannot be printed on demand, so the bank in turn must obtain it by liquidating and selling real projects.

But this is not how payments work, neither today nor in the past. To pay for consumption, you do not withdraw a suitcase full of cash and go to the store. Instead, in one manner or another, you direct your bank to credit that amount to the store. This is indeed one of the major reasons to use a bank in the first place, and has been possible in one form or another since long before there were credit cards or checking accounts.

Critically, banks do not need to shift any real resources to honor such requests. If your merchant uses the same bank as you, it can simply make two offsetting entries in a database. If your merchant uses a different bank, your

bank will need to transfer reserves held at the central bank to cover the transaction; but such transfers will typically be canceled out over time by transfers from the other bank's customers in the other direction. In the meantime, the bank can avoid liquidating real investments through the combination of netting, interbank borrowing, and central bank lending. Even before modern central banking, smaller-scale networks existed to accomplish similar goals.

This makes all the difference to the paper's story. With the payments technology described above, banks generally will *not* need to liquidate real investments in order to meet depositor demands. Indeed, this is a major function of the banking system in the first place. Then it is hard to see how bank runs can arise in a purely depositor-focused "panic" story, as described by the paper's stronger implication.[5]

Again, even if we do not accept that runs are driven *purely* by depositor psychology, we do not need to abandon the more general idea that strategic complementarities between depositors play a role. Think back to Silicon Valley Bank. The bank held a portfolio of long-term, fixed-rate debt with low credit risk (Treasuries, MBS, etc.). Everyone knew the bank was "long-run" solvent if left alone, since these securities would pay all their promised cash flows. But it was "short-run" insolvent in the sense that, if all depositors decided to withdraw, the market value of its securities would not be enough to honor the nominal value promised to those depositors. The bank floated along for a while without pressure, then collapsed suddenly.

This certainly seems to fit the more general idea of strategic complementarity between depositors, as in Diamond and Dybvig (1983). Indeed, the influence of the paper is such that many people will cite it to explain every bank run in practice. However, we must emphasize that the concern with SVB arose initially due to concerns about the bank's investment losses, *not* due to issues purely on the depositor side. In this sense, the events do not fit the paper's pure panic story. Again, most narratives of bank runs in practice sound more like this one, with a triggering event based on perceived losses by the bank.

Why is this important? If we believe the "panic" story in Diamond and Dybvig (1983), then major disruptions to the economy can occur through spontaneous and arbitrary panics occurring for no clear reason. This is an exciting story, and certainly justifies a very active role for policymakers in guiding investor psychology, while at the same time absolving bankers and regulators of any blame for stress to the banking system.

On the other hand, if bank runs are first triggered by fundamental and jus-

[5]Gorton and Winton (2003) discuss this issue as follows: "[In the DD model,] bank liabilities do not circulate as a medium of exchange. Instead, when a consumer learns that he has preferences for early consumption, he withdraws from the bank to satisfy those needs. There is no purchase of consumption goods using bank liabilities as money. In the model, the bank is, in effect, also the store. But, in cash-in-advance type models or search-theoretic models, consumers buy goods with bank liabilities without any need to return to the bank to withdraw. This is the essence of a medium of exchange. And that is how bank notes and bank deposits work. While consumption smoothing, and the demand for consumption insurance, are likely important features of reality, it is not clear that consumption smoothing is really a meaningful sense in which bank liabilities are a medium of exchange."

tified concerns about the bank's investment losses, then we might ask whether investor psychology is really so important after all. Perhaps we should just focus on having banks lose less money? This debate between the "panic" and "fundamental" view of bank runs has persisted to the present day. Most interestingly in my view, modern research has found ways to synthesize the two perspectives together, using techniques from the "global games" literature. See next section.

Global games to select equilibrium

A bank run in Diamond and Dybvig (1983) is interpreted as a sudden change between two equilibria. But, notice that the model is silent on how or when such a change could happen. The very idea of "switching" between equilibria raises questions that are difficult both philosophically and technically. The original paper, quite reasonably, does not attempt to dive into these issues.

However, the idea can be made rigorous with some additional machinery, specifically through a technique called "global games." The basic idea is to add a random fundamental into the model that causes the bank to be truly solvent or insolvent. Then we further add some noise in the observation of this fundamental, such that no one knows its value for sure. Finally, we endow each agent with a signal about that fundamental.

The outcome will be that for extreme values of the fundamental, everyone will correctly liquidate the bank or leave it alone. But for intermediate values, we again see "panic"-induced bank runs that are arbitrary but self-fulfilling: the bank may be liquidated even though it is solvent, due to the fear that *other* people got very bad signals. Then we can speak to the *probability* of bank runs and liquidation happening, and finally, can deduce conditions under which the banking setup is better or worse than other approaches to funding investment. The original techniques on global games were developed for other settings, most notably by Morris and Shin (1998) and Morris and Shin (2002). They were then applied to bank runs by Goldstein and Pauzner (2005) and Rochet and Vives (2004).

Global games are profoundly fascinating from both a technical and an economic perspective. Unfortunately, we will not have time to cover these methods in detail. However, with the high-level description just above, we can turn back to some of the questions posed in the prior section.

As I mentioned, in practice many people seem to describe bank runs as arising from "fundamental" shocks to the bank itself, not from pure panic by consumers (even though they still often explain the run by citing Diamond and Dybvig, 1983). The question becomes, are investor psychology and coordination still important considerations when bank runs are driven by such fundamentals?

The global-games papers provide a formal sense in which they are. In Goldstein and Pauzner (2005), the underlying banking model is the same as in Diamond and Dybvig (1983), but the bank also experiences fundamental shocks that depositors care about. Rochet and Vives (2004) go one step further and have shocks *only* to bank fundamentals, removing completely the shocks to depositor consumption needs. In both papers, the bank can experience a run, for

values of the fundamental where it is actually still solvent.

Thus we can rigorously say that the strategic complementarity between depositors can *amplify* the effect of fundamentals. In my view, Rochet and Vives (2004) is probably the closest model to what most people have in mind when explaining bank runs in practice today, even if they still typically cite Diamond and Dybvig (1983). Again, the big picture is that we can preserve the idea of strategic complementarity and multiple equilibria, which is really the main point of the paper, without needing to rely necessarily on their specific description of how the bank works or the idea of purely panic-driven runs, both of which are more questionable.

The structure of investment funds

We can also draw an analogy from the discussion above to many other settings in finance. One of the most important is the following:

Open-end funds holding illiquid assets, such as corporate bonds or real estate, can collapse easily during market downturns. The reason is that they promise investors the ability to redeem at NAV anytime. When the fund holds illiquid assets, the reported NAV is often based on stale prices, and at times may be clearly above the true market value of the assets. At these moments, there is a strong incentive for all investors to redeem as quickly as possible, out of fear that others will do the same and leave the fund with nothing. In response, the funds will often strategically suspend investors' redemption rights, to their great frustration. The logic is exactly like the bank facing a run. By contrast, closed-end funds and ETFs report NAVs, but never promise anyone the ability to redeem at that price. Hence, from a social perspective these fund structures are the natural way to hold illiquid assets.

An interesting exercise is to explain this last observation to investors, or the asset managers who create and market funds. In my experience, they typically give exactly the opposite opinion: Investors view the open-end structure as an advantage, *especially* when the fund holds illiquid assets, precisely because it allows them the ability to redeem anytime at NAV. They will typically suggest that NAV is the "correct" price, whereas the market price is somehow arbitrary and "wrong." They will acknowledge that fund collapses, or the suspension of redemptions, are terrible for investors when they happen. Yet they typically regard such events as rare enough to be ignored in practice.

I find this interesting because these market participants prefer run-prone vehicles, despite being much more sophisticated than the typical bank customer imagined in Diamond and Dybvig (1983). Perhaps this gives us some perspective on why these structures emerge over and over in the economy.

6.4 Conclusion

This chapter has provided only the smallest glimpse of research on banking theory. I have mainly tried to highlight the primary questions in this field,

and how those questions often end up applying in areas of finance that are not literally about the traditional notion of banks. However, there are many other seminal contributions to this literature that build on the insights presented here and offer new ones. (Some were even written by people other than Douglas Diamond!) Banking theory is really "financial intermediation theory," and every area of finance has reason to draw on the insights from this topic.

6.5 Exam question: Delegated monitoring

Consider the following modification of Diamond (1984).

- There are 8 risk-neutral depositors with one dollar each.
- There are two banks that can accept the depositors' funds and invest them in projects. However, the banks are capacity constrained, and can only accept four dollars of deposits each (and thus can fund two projects).
- Finally, there are four potential projects. Each project requires two dollars of initial investment to be provided to its manager. Two of the projects are in sector A and two are in sector B.
- Within a given sector, with probability p both projects generate $R > 1$ and with probability $1 - p$ both projects generate zero payoff. That is, payoffs are perfectly correlated within sectors, but independent across sectors.
- The model follows the "costly state verification" setup: When anyone in the model provides money to anyone else through an investment contract, they cannot directly observe the cash flow generated by the project unless they pay an auditing cost $\phi > 0$. If they do not pay the cost, they only observe a report from the funded party.

You can assume that all investment contracts in the model (between the depositor and bank, or between the bank and a project) resemble the optimal contract derived in the costly state verification framework of Diamond (1984): You do not need to prove the optimality of that contract. You can assume that a bank will always be able to repay depositors unless *both* of its projects fails.

There are two potential ways to set up the banking system in this model: Either each bank specializes in a sector (one bank funds both A projects and the other funds both B projects); or each bank diversifies across sectors (each bank funds one project from each sector). Identify which of these approaches has lower total expected auditing costs, and explain why. You do not need to give a formal proof, but give clear economic reasoning based on the model.

Solution to exam question 6.5

Remember that in Diamond (1984), the optimal contract looks like debt. The lender will pay the monitoring/auditing cost ϕ only in the state where the borrower does not repay.

Expected auditing costs *by the banks* are the same under either structure:

- Under the specialized structure, either both projects succeed or both projects fail. In the later case, both projects must be audited. Hence each bank's expected auditing cost is $(1-p) \times 2\phi$.
- Under the diversified structure, both of a bank's projects fail with probability $(1-p)^2$, and just one fails with probability $2p(1-p)$. So each bank's expected monitoring cost is $2p(1-p)\phi + (1-p)^2 2\phi = (1-p) \times 2\phi$.

Hence the banks' expected auditing cost is not a reason to favor either structure.

However, under the diversified structure, expected monitoring costs *by depositors* is lower. Under the specialized structure, each bank has probability $1-p$ of failing, so each depositor has expected monitoring cost of $(1-p)\phi$. Under the diversified structure, each bank has probability $(1-p)^2$ of failing, so each depositor has expected monitoring cost of $(1-p)^2\phi$.

Taking a step back, this highlights the benefit of diversification in the Diamond (1984) framework. That diversification is not to address risk aversion by depositors, but rather to lower the probability of any specific bank failing which would require depositors to incur monitoring costs.

Bibliography

George Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

Eduardo Azevedo and Daniel Gottlieb. Perfect competition in markets with adverse selection. *Econometrica*, 85(1):67–105, 2017.

Malcolm Baker and Jeffrey Wurgler. Market timing and capital structure. *Journal Finance*, 57:1–32, 2002.

Jeffrey Banks and Joel Sobel. Equilibrium selection in signaling games. *Econometrica*, 55(3):647–661, 1987.

Elazar Berkovitch and E. Han Kim. Financial contracting and leverage induced over- and under-investment incentives. *Journal of Finance*, 45(3):765–794, 1990.

Adolf Berle and Gardiner Means. *The Modern Corporation and Private Property*. The Macmillan Company, New York, NY, 1932.

Marianne Bertrand and Sendhil Mullainathan. Enjoying the quiet life? corporate governance and managerial preferences. *Journal of Political Economy*, 111(5):1043–1075, 2003.

David Besanko and Anjan V Thakor. Collateral and rationing: Sorting equilibria in monopolistic and competitive credit markets. *International Economic Review*, 28(3):671–689, 1987.

Helmut Bester. Screening vs rationing in credit markets with imperfect information. *The American Economic Review*, 75(4):850–855, 1985.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Michael Brennan and Alan Kraus. Efficient financing under asymmetric information. *The Journal of Finance*, 42(5):1225–1243, 1987.

Michael Brennan and Eduardo Schwartz. Optimal financial policy and firm valuation. *Journal of Finance*, 39(3):593–607, 1984.

Chun Chang. Payout policy, capital structure, and compensation contracts when managers value control. *The Review of Financial Studies*, 6(4):911–933, 1993.

Pierre-André Chiappori and Bernard Salanié. Testing for asymmetric information in insurance markets. *Journal of Political Economy*, 108(1):56–78, 2000.

Gabriella Chiesa. Debt and warrrants: Agency problems and mechanism design. *Journal of Financial Intermediation*, 2(3):237–254, 1992.

In-Koo Cho and David M Kreps. Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 102:179–221, 1987.

Russell Cooper and João Ejarque. Financial frictions and investment: Requiem in *Q*. *Review of Economic Dynamics*, 6:710–728, 2003.

Gregory Crawford, Nicola Pavanini, and Fabiano Schivardi. Asymmetric information and imperfect competition in lending markts. *American Economic Reivew*, 108(7):1659–1701, 2018.

Partha Dasgupta, Peter Hammond, and Eric Maskin. The implementation of social choice rules. *The Review of Economic Studies*, 46(2):185–216, 1979.

David de Meza and David C Webb. Too much investment: A problem of asymmetric information. *The Quarterly Journal of Economics*, 102(2):281–292, 1987.

Douglas Diamond. Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51:393–414, 1984.

Douglas Diamond and Philip Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419, 1983.

Douglas W Diamond and Raghuram G Rajan. Liquidity risk, liquidity creation, and financial fragility: A theory of banking. *Journal of Political Economy*, 109(2):287–327, 2001.

Pradeep Dubey, John Geanakoplos, and Martin Shubik. Default and punishment in general equilibrium. *Econometrica*, 73(1):1–37, 2005.

Liran Einav and Amy Finkelstein. Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, 25(1):115–138, 2011.

Timothy Erickson and Toni M Whited. Measurement error and the relationship between investment and q. *Journal of Political Economy*, 108(5):1027–1057, 2000.

Douglas Gale and Martin Hellwig. Incentive-compatible debt contracts: The one-period problem. *Review of Economic Studies*, 52:647–663, 1985.

Ronald Giammarino and Tracy Lewis. A theory of negotiated equity financing. *The Review of Financial Studies*, 1(3):265–288, 1987.

Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41:587–601, 1973.

Itay Goldstein and Ady Pauzner. Demand-deposit contracts and the probability of bank runs. *Journal of Finance*, 60(3):1293–1327, 2005.

João F Gomes. Financing investment. *American Economic Review*, 91(5):1263–1285, 2001.

Gary Gorton and George Pennacchi. Financial intermediaries and liquidity creation. *Journal of Finance*, 45(1):49–71, 1990.

Gary Gorton and Andrew Winton. Financial intermediation. volume 1, pages 431–552. 2003.

John Graham. How big are the tax benefits of debt? *The Journal of Finance*, 55(5):1901–1941, 2000.

Jerry Green and Jean-Jacques Laffont. Characterization of satisfactory mechanisms for the revelation of preferences. *Econometrica*, 45:427–438, 1977.

Milton Harris and Artur Raviv. Optimal incentive contracts with imperfect information. *Journal of Economic Theory*, 20:231–259, 1979.

Oliver Hart and John Moore. Debt and seniority: An analysis of the role of hard claims in constraining management. *The American Economic Review*, 85(3):567–585, 1995.

Fumio Hayashi. Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica*, 50(1):213–224, 1982.

Martin F Hellwig. A reconsideration of the Jensen-Meckling model of outside finance. *Journal of Financial Intermediation*, 18:495–525, 2009.

Christopher Hennessy and Toni M Whited. Debt dynamics. *Journal of Finance*, 60(3):1129–1165, 2005.

Robert Innes. Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory*, 52:45–67, 1990.

Michael C Jensen and William H Meckling. Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, 3:305–360, 1976.

David Kreps and Robert Wilson. Sequential equilibria. *Econometrica*, 50:863–894, 1982.

Owen Lamont. Corporate-debt overhand and macroeconomic expectations. *The American Economic Review*, 85(5):1106–1117, 1995.

Hayne Leland. Agency costs, risk management, and capital structure. *Journal of Finance*, 53(4):1213–1243, 1998.

Hayne E Leland and David H Pyle. Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance*, 32(2):371–387, 1977.

Neale Mahoney and E. Glen Weyl. Imperfect competition in selection markets. *The Review of Economics and Statistics*, 99(4):637–651, 2017.

Robert C Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2):449–470, 1974.

Paul Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391, 1981.

Merton Miller. Debt and taxes. *The Journal of Finance*, 32(2):261–275, 1977.

Merton Miller and Kevin Rock. Dividend policy under asymmetric information. *The Journal of Finance*, 40(4):1031–1051, 1985.

Franco Modigliani and Merton H Miller. The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48(3):261–297, 1958.

Franco Modigliani and Merton H Miller. Corporate income taxes and the cost of capital: A correction. *The American Economic Review*, 53(3):433–443, 1963.

Dilip Mookherjee and Ivan P'ng. Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics*, 104(2):399–415, 1989.

Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, 88(3):587–597, 1998.

Stephen Morris and Hyun Song Shin. Social value of public information. *American Economic Review*, 92(5):1521–1534, 2002.

Stewart C Myers. Determinants of corporate borrowing. *Journal of Financial Economics*, 5:147–175, 1977.

Stewart C Myers. Still searching for optimal capital structure. *Journal of Applied Corporate Finance*, 6:4–14, 1993.

Stewart C Myers and Nicholas S Majluf. Corporate financing and investment decisions when firms have information that investors do not have. *The Journal of Financial Economics*, 13:187–221, 1984.

Roger Myerson. Incentive compatibility and the bargaining problem. *Econometrica*, 47(1):61–73, 1979.

David C Nachman and Thomas H. Noe. Optimal design of securities under asymmetric information. *The Review of Financial Studies*, 7(1):1–44, 1994.

Thomas H. Noe. Capital structure and signaling equilibria. *The Review of Financial Studies*, 1(4):331–355, 1988.

Robert Parrino and Michael S Weisbach. Measuring investment distortions arising from stockholder-bondholder conflicts. *Journal of Financial Economics*, 53:3–42, 1999.

Rafael Repullo and Javier Suarez. Monitoring, liquidation, and security design. *The Review of Financial Studies*, 11(1):163–187, 1998.

Jean-Charles Rochet and Xavier Vives. Coordination failures and the lender of last resort: Was bagehot right after all? *Journal of the European Economic Association*, pages 1116–1147, 2004.

Stephen A Ross. The determination of financial structure: The incentive-signalling approach. *The Bell Journal of Economics*, 8(1):23–40, 1977.

Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649, 1976.

Steven Shavell. Risk sharing and incentives in the principal and agent relationship. *The Bell Journal of Economics*, 10(1):55–73, 1979.

Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87:355–374, 1973.

Joseph E Stiglitz and Andrew Weiss. Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3):393–410, 1981.

Jean Tirole. *The Theory of Corporate Finance*. Princeton University Press, Princeton, NJ, 2006.

Robert M Townsend. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21:265–293, 1979.