Investment & Portfolio Decisions with Uncertainty and Market Frictions: Theory and Application to Microfinance

Dissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät

der Universität Regensburg

vorgelegt von

Johannes J. Reeder

Berichterstatter:

Prof. Dr. Lutz Arnold (Universität Regensburg)

Prof. Sven Rady, Ph.D. (Ludwig-Maximilians-Universität München)

Tag der Disputation: 12. April 2010

Contents

List of Figures

List of Tables

Preface

This dissertation is the result of three years of research in the economics department at the Chair of Economic Theory at the University of Regensburg. During that time, I held a scholarship of the Bavarian Graduate Program in Economics (BGPE). When I faced the opportunity to join the program, I did not have to balance reasons since prospects were so appealing. The BGPE aims to foster research cooperations within and between economics departments of Bavarian universities. Apart from supervision by two economics professors from two different Bavarian universities, the means to do so include courses, research workshops and extra-curricular activities. The courses cover general topics in macroeconomics, microeconomics and econometrics, as well as special subject courses. They are open to all Bavarian doctoral students in economics and have proved to be a source of numerous fruitful discussions. Twice a year, the research workshops provide the unique opportunity to get substantial feedback from both university professors and doctoral students from all Bavarian universities. In addition to academic cooperations within Bavaria, the BGPE tries to boost international affiliation by inviting international guest researchers to give talks at Bavarian universities. Moreover, being free from teaching obligations, BGPE scholars stand the chance to do research abroad.

During the last three years, I have had both good and more difficult times. I would like to express my gratitude to Prof. Dr. Lutz Arnold, who is one of the founding fathers of the BGPE and the first supervisor of my thesis. He was always ready for discussion, provided me with detailed comments, and gave me a fresh research impetus every time I was in doubt about my work. I also benefitted from the support of my second supervisor, Prof. Sven Rady, Ph.D., who motivated me at the beginning of my thesis and encouraged me to stay the course toward the end. Prof. Regina Riphahn, Ph.D., as another BGPE founding member and spokesperson of the program must be assigned responsibility for many of the aforementioned BGPE features I stood to benefit from. During my research stay at Simon Fraser University (SFU) in Vancouver, I was able to expedite a project on microfinance, which is the subject of Chapter 3. I would like to thank the academic staff of SFU, who helped me to get a broader view on the topic. In particular, I am indebted to Prof. Christoph Luelfesmann, who invited me to stay at SFU and gave me numerous valuable comments in the process of the project, and to Alex Karaivanov as another source of inspiration.

I would also like to thank all teaching assistants and professors at the University of Regensburg for general discussions, comments and research cooperations. Most of this dissertation is based on three joint papers. The first is joint work with Stefanie Trepl, the second is done together with Prof. Dr. Lutz Arnold and Susanne Steger. The third paper stems from a research cooperation with Prof. Dr. Gregor Dorfleitner and Michaela Leidl from the Center of Finance. A constant source of valuable information regarding statistical concepts was Kilian Plank. I have greatly benefited from discussions with all the aforementioned colleagues.

Also, I have to thank my parents, Ursula and Frank Reeder, and my sister Ulrike Reeder for advice and assistance during the past three years, and my friends who have shown understanding for little time for them.

Chapter 1

Introduction

According to Eichberger and Harper (1997, p.xi), "...a good grounding in microeconomic theory is considered essential to a proper grasp of the principles of finance". Financial economics combines insights from the literature on finance and economics using theoretical models. Typically, these models consider decisions at different points in time and allude to the concepts of uncertainty and risk. This dissertation presents three distinct theoretical models which complement the existing theory in the respective area of financial economics.

Investment behavior and portfolio choice are similar topics. Both relate to the idea of dispensing with current consumption for the benefit of future profit. Several actors in an economy face this decision. Households consisting of individuals have to schedule their consumption behavior. In order to do a trip around the world next year, today's consumption must be reduced. Firms are subject to demand constraints and must decide whether to strengthen their operative units today or to keep their funds safe to invest in the future. However, 'investment behavior' is frequently used to describe the process of 'how much', whereas 'portfolio choice' is more concerned with the 'how'. Applying this distinction, Chapter 2 is more closely related to investment behavior. Firms face an investment decision and have to get funding from banks. Banks have to obtain refinancing from households by offering them interest rates on deposits, which in turn depend on how much and which firms choose to invest. Chapter 3 focuses on microfinance institutions which obtain funding from capital markets and decide subject to market forces how to channel funds to micro-entrepreneurs. Chapter 4 considers a classical portfolio choice problem in which a given amount of money is to be allocated between a given set of securities.

Households' consumption-savings and firms' investment decisions are directly connected to uncertainty and risk. Buying a car today gives immediate pleasure. By contrast, money invested in some security might become less valuable due to price fluctuations, or might even be completely lost if the issuer of the security declares default. Depending on preferences in general and risk attitude in particular, consumption-savings decisions will differ between households and so will investment plans between firms. A common feature of the decisions of both households and firms involves some kind of optimization. The theory of decisions under uncertainty has proposed several ways to represent and model decision making. The models in Chapters 2 and 3 assume that households (firms) take decisions in order to maximize expected utility (profits), but Chapter 2 also extends to consequences of non-expected utility maximization. In Chapter 4, we set up a model which assumes that investors behave in order to maximize some function of statistical moments, similar to the standard mean-variance approach.

In Chapters 2 and 3 we focus on a particular type of market friction and its respective implications for equilibrium outcomes. Even though the idealized concept of perfect markets serves to analyze basic relationships in an economy, it is highly fictitious and unable to help explain many of the phenomena observed in our world.

A market friction which has received considerable attention not only in the realm of financial markets is asymmetric information. The literature started to incorporate the fact that information is neither perfect nor symmetric some decades ago.[1] Stiglitz (2002, p.461) argues that researchers focusing on perfect information models were aware of the fact that information is imperfect, but that the academic climate of the era was to hope that results for markets with minor information asymmetries were similar to results obtained assuming symmetric information.

Much of this research has taken place in the realm of contract theory analyzing principals contracting with agents. Hart and Holmström (1987) speak of adverse selection models when agents have precontractual information. By contrast, in moral hazard models, information is symmetrically distributed at the time of contracting. A further distinction can be made: If

[1] For a survey on different types of asymmetric information, see Laffont and Martimort (2002, Ch.1). A description of the (change of the) role of information in economic theory during the last decades is given by Stiglitz (2000, 2002).

there is asymmetric information in the contract period, moral hazard is said to obtain 'ex ante'. Otherwise, i.e., if asymmetric information exists only after the duration of the contract, there is 'ex post' moral hazard. Classical examples for ex ante moral hazard are, first, the employee whose effort at work is imperfectly observable and who might thus shirk and, second, the insuree who undertakes more risky actions than if uninsured. In the realm of credit markets, the two examples are also present: Borrowers might be lazy and thus jeopardize contractual repayment, or they might choose to invest the money in a project which goes against the bank's interest, e.g., with limited liability, a project with high risk. Ex post moral hazard is typical of credit markets. If the bank cannot observe how successful borrowers are, the latter might simply tell the bank a lie about their revenues. Also, borrowers might not consider to repay but 'take the money and run' instead. Armendáriz de Aghion and Morduch (2005) describe this latter action both as ex post moral hazard and as an enforcement problem. In fact, the distinction between the latter two concepts is blurred, but we propose to interpret the 'take the money and run' phenomenon as a problem arising from imperfect enforcement.[2]

Assuming perfect enforcement of contracts is not as critical as the assumption of symmetric information, but very inadequate in some specific contexts. If the analysis is to explain a phenomenon in a developed country, enforcement is indeed a minor problem. A sound legal system along with a corruption-free executive makes perfect enforcement a good approximation to reality. The most important obstacle to perfect enforcement is then debtor protection by limited liability. By contrast, the world consists to a large extent of developing and transition economies, which are frequently characterized by institutional problems, e.g., poor property rights or high degrees of corruption. If theoretical models are to properly reflect economic choices of agents in such economies, it is highly doubtful to assume perfect enforcement. In Chapter 3, we consider a model of credit markets in developing countries, where enforcement problems are a major impediment to welfare-enhancing trade.

As mentioned above, this dissertation contributes to the existing theory of financial markets. Chapter 2 reexamines one of the most influential models in this area, viz., Stiglitz and Weiss (1981) (SW, henceforth). In Chapter 3, we conduct an equilibrium analysis of the

[2]Note that the classification of the 'take the money and run' phenomenon depends on the kind of money referred to. If a borrower takes the funds obtained to invest and then runs, we have a situation of ex ante moral hazard. By contrast, borrowers taking the proceeds of their investments and running matches the ex post moral hazard or enforcement problem definitions.

Besley and Coate (1995) (BC, henceforth) model to describe credit market outcomes in the realm of microfinance. Chapter 4 extends the theory of portfolio choice by adding a social dimension to the classical mean-variance model of Markowitz (1952).

The first model, presented in Chapter 2, contributes to the literature on 'equilibrium credit rationing'. In light of the forthcoming publication of Arnold and Riley (2009) (AR, henceforth), the results of the seminal model of SW require closer scrutiny. A particularly important issue is the consideration of dependency of firms' revenues and the analysis of consequences for equilibrium. In SW and AR, revenues are assumed to be independent. This assumption seems rather restrictive given the global economic ties and the recent correlated movement of asset prices. The additional assumption of a large number of firms in SW and AR leads to perfect diversification of (idiosyncratic) risk and, hence, a riskless deposit rate that can be passed through to suppliers of capital. Therefore, SW and AR model capital supply in a highly simplified way. The consideration of dependent revenues calls for explicit modeling of households' consumption-savings decision since the deposit rate passed through to households is not riskless in that case.

We show that the type of equilibrium can crucially depend on the degree of dependence of project revenues. Capital risk deters households from saving so that there might be a credit rationing equilibrium. Defining the social optimum, we find that project dependence might reduce the number of safe projects carried out in equilibrium in a socially harmful way. Thus, project dependence can aggravate adverse selection. In three extensions, we show how risk aversion, imperfect revenue dependence and a different structure of dependence influence the results. Our analysis points out that project dependence is an important factor in the determination of credit market outcomes.

Chapter 3 is based on the observation that many microfinance institutions (MFIs) approach financial self-sufficiency, which improves their ability to compete for funds on the capital market. At the same time, the use of market instruments increases. This brings up the question of what market equilibria in microfinance markets look like and which kinds of market failure tend to afflict them. Our starting point is the seminal model of BC, who put a game-theoretic structure on the repayment behavior of borrowers under joint liability. We compare standard 'individual lending' (IL) to 'group lending' (GL). One result is that the repayment rate comparison of BC is not sufficient to predict market outcomes, as it is biased toward group lending. The market outcomes with non-cooperative repayment behavior of group members are compared to the results under the assumption of cooperative behavior. A characterization of market equilibria shows that microfinance markets can suffer from market failures known from the adverse selection literature, namely financial fragility, redlining, and credit rationing. Social sanctions ameliorate these problems, but do not eliminate them.

In Chapter 4, we challenge the frequently made assumption that economic agents act upon purely materialistic grounds, i.e., that decisions are taken only to maximize (utility derived from) income and wealth. Even though there seems to be a consensus that it is not the only decision criterion, most theoretical models use it as a workhorse. We complement standard portfolio theory à la Markowitz by adding a social dimension. We distinguish between two main setups, taking social returns as stochastic in the first, but as deterministic in the second. Two main features need to be introduced: Every asset must be assigned a (distribution of) social return(s), and the investor has to cherish social returns. The former is subject to measurement problems, whereas the latter is mainly a problem of choosing a suitable utility representation. The main result involves the existence of a unique optimal portfolio of risky assets for all investors, as in Tobin (1958). If there is a riskless asset, we show that different types of investors usually have different optimal portfolios of risky assets. Interestingly, if investors differ in risk aversion only, there is a unique optimal portfolio of risky assets, and only the shares of wealth invested in the riskless asset and that portfolio differ.

Each chapter starts with a section in which we lay out in detail why we study the respective topic. These sections also provide a survey of the respective literature. Moreover, each chapter has its own appendix with additional material like numerical calculations or proofs. In the main text, we present those proofs which are either constructive or theoretically demanding, whereas pure algebra is delegated to the appendix. Chapter 5 concludes.

Chapter 2

Asymmetric Information in Credit Markets

This chapter is based on joint work with Stefanie Trepl. It presents an extended version of Reeder and Trepl (2009).

2.1 Motivation and the literature

From business cycle theory, we know that most economic variables fluctuate over time. During expansive periods, aggregate economic activity increases and so do aggregate profits. In a recession, the opposite occurs. Ups and downs over time indicate that firms' profits are not independent of each other, but depend on common economic factors. The possible extent of these interdependencies has been highlighted by the recent financial crisis, which caused most industries serious financial distress. Households lost a fortune, not only through highly risky investments. Given that firms' profits are highly dependent[1] and aggregate risk is enormous, how does this influence savings of households? We try to answer the question of whether households' awareness of risk might make them reduce savings such that firms do not obtain the funds they would like to borrow. Asymmetric information plays a major role in answering that question. Since firms know more about the risk characteristics of their projects than households and banks do,[2] the problem of high aggregate risk cannot be avoided by only funding safe firms. Aggregate risk can have tremendous consequences for one of the banks' most important tasks, diversification. Revenue dependence within a single bank's credit portfolio is one of the main research topics in financial risk management. However, to our best knowledge, there is no literature that analyzes the consequences of dependence of firms' revenues for credit market equilibria in an adverse selection model.

In banking theory, credit rationing is an important phenomenon. Bhattacharya and Thakor (1993) mention six fundamental puzzles in financial intermediation research. The second is about allocation of credit and questions "why banks deny credit to some rather than charging higher prices" (p.3). There is a vast literature trying to answer this question.

In the very early literature, the term credit rationing was mainly used in connection with the effects of monetary policy, an example being Rosa (1951)[3]. Scott (1957) assumed that banks hold government debt and private sector loans at the same time and react to changes in monetary policy with redeployment of capital between the two. Other advocates of the so-called 'availability doctrine' stressed the effects of monetary policy not only on the supply but

[1] Instead of 'dependence', some readers might prefer to speak of 'correlation'.

[2] Almost all of the literature on asymmetric information assumes that financial intermediaries are less well informed about projects of firms. However, Hillier and Ibrahimo (1993, p.300) point to the possibility of a converse information structure citing young firms as a convincing example.

[3] For an explication of the confusion regarding the spelling of the author's last name, see footnote 3 on page 273 in Hillier and Ibrahimo (1993).

on the credit demand side, too. Uncertainty and changing expectations of market participants were frequently used explanations: If firms expect that an increase of the federal funds rate will be followed by a decrease, they reduce current demand for funds since conditions in the future might become better. A comprehensive description of the numerous contributions around the availability doctrine is given in Jaffee (1971, Ch.2). Notwithstanding widespread criticism, Baltensperger (1978, p.170) attributes great importance to the doctrine, namely that it "suggested an alternative transmission channel for monetary policy that was ... based on a credit rationing argument".

A unanimous definition of credit rationing is missing to this day. It is helpful to distinguish credit rationing from general price rationing in economics. If the Walrasian market price is too high for someone to take a loan, he is frequently said to be rationed by the market mechanism. This is not what credit rationing means in the literature. Instead, credit rationing is said to occur if a borrower is ready to pay the market interest rate, but does not get credit nonetheless. However, credit contracts do not only stipulate an interest rate. Section 5.1 in Freixas and Rochet (2008) refers to Baltensperger (1978), defining "*equilibrium credit rationing* as occurring whenever *some borrower's demand for credit is turned down, even if this borrower is willing to pay all the price and nonprice elements of the loan contract*" (Freixas and Rochet, 2008, p.172). The loan rate is the most obvious price element, examples for nonprice elements are the amount of collateral, loan size or maturity of the loan.

Most of the contributions in the field use very specific models. Based on the respective assumptions, the authors propose definitions which allow them to establish the (non-) existence of credit rationing in their respective setups. A helpful distinction has been proposed by Keeton (1979). On the one hand, "credit is rationed whenever a customer receives a loan of smaller size than he would desire at the interest rate quoted by the bank" (p.2). On the other hand, there is credit rationing "when some firms are able to obtain loans while other, identical firms are not" (p.2). The literature has referred to these two concepts as rationing of 'type I' and 'type II', respectively. According to Elsas and Krahnen (2004, p.216), "[c]redit rationing is an economic phenomenon typically associated with problems of information asymmetries or incomplete contracting in debt markets". This definition focusses on the more recent publications which establish credit rationing as arising endogenously due to some plausible market friction, like the ones mentioned in the definition. In contrast, as pointed out by Clemenz

(1986, p.3), the early approach to credit rationing assumed rather than explained it. As an example, Clemenz mentions ad hoc rigidities which - rather unsurprisingly - lead to credit rationing by the standard mechanisms of supply and demand theory. Such ad hoc rigidities frequently consist of interest rate regulations as, for instance, the existence of usury laws.[4]

It was in the late seventies that the literature started to incorporate the fact that credit market participants do not necessarily share the same information, and to abandon the assumption of complete state-contingent contracting. Critics argued that there was no way to explain credit rationing in consistency with rational, profit-maximizing lenders, taking into account both credit supply and demand. SW resorted to research following Akerlof (1970)[5] and introduced asymmetric information as a crucial assumption in models of the credit market. They were - supposedly[6] - able to explain how adverse selection resulting from asymmetric information might lead to an equilibrium with credit rationing. In SW, the term credit rationing is used "...for circumstances in which either a) among loan applicants who appear to be identical some receive a loan and others do not, and the rejected applicants would not receive a loan even if they offered to pay a higher interest rate; or b) there are identifiable groups of individuals in the population who, with a given supply of credit, are unable to obtain loans at any interest rate, even though with a larger supply of credit, they would" (pp.394-395). SW argued that an increase of the loan rate might decrease banks' expected returns since some 'good' borrowers do not demand credit any more, notwithstanding the remaining borrowers paying a higher rate (if they are able to pay back). The equilibrium in their model entails credit rationing if the decrease is such that the return function is 'globally hump-shaped'.

As in SW, most of the theory on equilibrium credit rationing is based on the possibility of a backward-bending credit supply curve: From a certain loan rate on, an increase in the loan rate could make lenders reduce their supply of credit due to a decrease in the return on lending. In SW, this follows from the fact that banks' returns can - supposedly - be globally hump-shaped. Another line of argument explains a backward-bending credit supply by resorting to changes in default risk due to decreasing returns to scale of projects with variable loan size. Baltensperger (1978, p.171) summarizes the argument as "showing that,

[4]In a standard supply and demand diagram, it can easily be seen how a usury law stipulating a maximal interest rate below the market-clearing rate leads to excess demand at the former rate.

[5]Using the market for used cars, Akerlof first explained how asymmetric (quality) information can prevent that markets clear.

[6]We come back to this point later in this section.

after a certain loan size is reached, *no* increase in the rate of interest can compensate the lender for the increased default risk associated with further increases in the size of the loan".[7] Both the SW approach and the default risk explanation lead to a backward-bending supply curve for the same reason, viz., the dependence of the loan portfolio's quality on the interest rate. The direct gains from charging a higher loan rate might be offset by a reduction in the (average) quality of the remaining borrowers.[8]

Asymmetric information has been modelled in various ways (cf. Chapter 1). In SW, two types of ex ante asymmetric information are shown to possibly cause credit rationing: hidden information and hidden actions. If there is hidden information, banks face different types of firms which they cannot distinguish when they decide about which firms to fund. If borrowers can take actions which influence the project payoff (or the probabilities of payoffs) after a loan has been granted, firms might be tempted to commit 'moral hazard', i.e., to take actions unobservable for (hidden from) the bank which are good for them at the expense of the bank.[9] Williamson (1987) considers ex post asymmetric information referring to moral hazard after project returns have been realized. The central claim is that banks cannot observe project revenues of firms without a cost.[10]

Even though widely recognized as a seminal contribution to the literature, many of the assumptions in the SW paper have been criticized and modified to yield a variety of interesting results. Riley (1987) assumes that banks are able to classify borrowers into risk classes. He shows that credit rationing cannot occur in more than one of these classes and concludes that "the extent of rationing generated by the S-W model is not likely to be empirically important" (p.224). The neglect of other markets, the equity market in particular, elicited work on the interaction of equity and debt financing. Hellmann and Stiglitz (2000) consider the mutual compatibility of debt and equity markets, as well as their interaction. They also provide an excellent survey of the literature of rationing under asymmetric information. De Meza

[7] Given the assumptions underlying this result, Baltensperger does not consider it as surprising.

[8] In Chapter 3, we will see how credit rationing can occur even if borrowers are (ex ante) homogeneous and information is symmetric. Due to enforcement problems, higher loan rates make strategic default more probable so that returns decrease.

[9] Other important contributions emphasizing hidden actions are Jaffee and Russell (1976) and Bester and Hellwig (1987).

[10] Williamson adapts the model of Gale and Hellwig (1985) to point out how monitoring costs can imply optimality of the standard debt contract à la Townsend (1979). Furthermore, in Williamson (1986), he is able to show how bank intermediation arises endogenously: Financial intermediaries economize on monitoring costs. In both of his papers, credit rationing can exist in equilibrium.

and Webb (1987) analyze how the method of finance, debt or equity, endogenously arises in equilibrium. They compare the equilibrium investment level to the socially optimal level and show that the SW assumptions lead to underinvestment, whereas their own assumptions imply overinvestment in equilibrium. The optimum method of finance differs between both situations: equity in the SW setup, but debt in their own setup. The results depend on the assumptions which differ, first, in terms of the asymmetry of information. Whereas lenders are uninformed about the expected return of a potential borrower in the model of de Meza and Webb, it is only the risk of a borrower's project which lenders cannot observe in the SW model.[11] Second, de Meza and Webb's analysis tackles another critical feature of the SW setup, namely the specific distribution of project returns (cf. Section 5.1 in Hillier and Ibrahimo, 1993). In SW, all projects are assumed to yield the same expected return and, thus, to differ only in risk. In the model of de Meza and Webb, firms have the same return if successful but different success probabilities. Thus, safer projects have higher expected returns. It is an empirical question which of the two setups is more adequate.[12]

It has also been criticized that SW assume the use of standard debt contracts. Most importantly, this excludes that contracts can be used as a sorting mechanism in order to induce self-selection of borrowers. In contrast to SW, who assume an exogenous amount of collateral, Bester (1985) allows banks to offer various loan contracts specifying loan rate and amount of collateral at the same time to show that, if an equilibrium exists, it is not characterized by credit rationing. In his model, less risky firms choose the contract with higher collateral requirement at a lower loan rate, and vice versa. Another sorting mechanism is loan size. Its analysis requires departing from the SW assumption of fixed capital requirement. In Milde and Riley (1988), banks use large loans at high rates which attract less risky borrowers. Models in which banks induce self-selection of borrowers can also be found in Chan and Kanatas (1985), who discuss different types of collateral, and Besanko and Thakor (1987b), where the focus is on the influence of market structure on credit allocation.

Another critical remark about the SW model concerns the equilibrium concept. There is no game-theoretic foundation which would enable us to establish a market equilibrium as

[11] An attempt to clarify the differences between the two setups is made by Bernhardt (2000). He shows how differences in both the algebraic formulation of and the kind of uncertainty about production technology can explain the over- and underinvestment results.

[12] We think that the assumption of de Meza and Webb is less realistic than the SW proposal since there is an ex ante dominance between projects in the model of de Meza and Webb.

the equilibrium of a game between banks. Freixas and Rochet (2008, p.174) note that "the implicit rules of the game are that banks are price setters on the credit market and quantity setters on the deposit market". An analysis of intermediation in a more general framework is given by Stahl (1988). He applies strict game-theoretic reasoning to the behavior of merchants who mediate between suppliers and consumers. One of the main results is that the order of moves in the game is crucial for the existence of a Walrasian equilibrium. If we interpret banks as merchants that sell loans to firms and compete for households' deposits, Stahl's result implies that there is a unique Walrasian subgame perfect Nash equilibrium if the credit subgame precedes the deposit subgame.[13]

Critics have also argued that the bank-borrower relationship is more complex than it can be expressed in the static setup of SW. As a reply, Stiglitz and Weiss (1983) introduce multiperiod relationships to show that credit rationing is still possible. Diamond (1989) considers the incentive effects of borrower reputation in a dynamic setup. Focussing on the role of collateral, Bester (1994) analyzes the effects of debt renegotiation on the design of optimal credit contracts given asymmetric information. Lending relationships are also considered by Petersen and Rajan (1995), who point out the importance of competition. De Meza and Webb (2006) allow firms to postpone the realization of projects so that firms can influence debt capital requirements by accumulating capital resources over time.[14] They conclude that credit rationing only appears under very specific conditions.

By now, the reader might have noticed that, even in such a specific area as equilibrium credit rationing given asymmetric information, there are many seemingly contradictory results. For instance, some models show that credit rationing can occur in equilibrium, whereas others show the contrary. If credit rationing can occur, it might but need not be socially inefficient, depending on the assumptions of the model. Even though it looks unsatisfactory at first glance, this is typical of models in the realm of information economics. We concur with Clemenz (1986, p.199) who notes that "there is only one way of being perfectly informed, but a myriad of possibilities for information to be incomplete".

[13]Based on the papers of Stahl (1988) and Yanelle (1989, 1997), who pioneered double Bertrand competition, Arnold (2007) applies a rigorous game-theoretic equilibrium concept to the SW model. Given some weak assumptions on the shape of capital supply and demand, he shows that the two-price equilibrium occurs in any subgame perfect pure-strategy equilibrium if the credit game precedes the deposit game.

[14]De Meza and Webb also consider other means to influence debt capital requirements, as for instance the downscaling of projects.

The results of SW have many implications.[15] Among others, the 'Law of Supply and Demand' and, therefore, standard comparative statics analysis breaks down. Supply and demand are found to be interdependent and the 'Law of the Single Price' is not valid any more (cf. the description of a two-price equilibrium in SW and in Subsection 2.2.5 of this chapter). In terms of welfare, asymmetric information might cause inefficient levels of investment[16] so that there might be a case for government intervention. An interesting paper on the role of governments as lenders of last resort is Mankiw (1986), who finds that the government can improve on market allocations. However, as noted by Hillier and Ibrahimo (1993, p.288), "it is dangerous to make strong policy recommendations on the basis of such a simple model".[17] Even though their statement refers to the model of de Meza and Webb, we think it is valid for most of the models in the domain of credit rationing.

Theory argues that imperfections on capital markets, such as credit rationing, can have serious macroeconomic consequences. In a discussion of the papers in Part VII of Bhattacharya, Boot, and Thakor (2004), Reichlin (2004) emphasizes the importance of financing activities, financial market access and the choice of contractual arrangements to a strand of literature which he calls the "financial structure approach to macroeconomics" (p.856). In particular, economic growth and development can be crucially affected by the functioning of financial markets. There are many papers and textbooks on this topic, an excellent survey is given by Goodhart (2004). Another line of research considers the influence of financial market arrangements on the business cycle, an example being Greenwald and Stiglitz (1993).

The variety of results mentioned above shows the economic significance of asymmetric information. The underlying assumptions crucially affect equilibrium outcomes in theoretical models. However, the economic significance of credit rationing should be empirically validated. Unfortunately, empirical evidence is scarce due to obvious data restrictions. It is hard to obtain micro data on the contractual terms of commercial bank loans. Macroeconomic proxies used include the speed of commercial loan rate convergence,[18] and the drawdown of trade

[15] For a comprehensive description of the implications of the SW model, the reader is referred to Section 4 in Hillier and Ibrahimo (1993).

[16] We analyze this point in Subsection 2.2.5.

[17] We know from De Meza and Webb (1992) that rationing need not be inefficient.

[18] If an increase in open-market rates is not (or only with delay) followed by increasing commercial loan rates, loan rates are said to be 'sticky'. This is taken as evidence for credit rationing, even though theory suggests many other explications of sticky loan rates. Two such studies (with opposing conclusions) using loan rate stickiness as a proxy are Jaffee (1971) and Slovin and Sushka (1983).

credits as in Petersen and Rajan (1997). An exception is the paper of Berger and Udell (1992), which uses a large micro data set from the Federal Reserve System. On the one hand, they cannot exclude the possibility of information-based credit rationing in equilibrium. However, they also state that their results do not give support for the hypothesis that information-based equilibrium credit rationing is an important macroeconomic phenomenon. Clearly, their findings depend on the design of the investigation and the specific data set. Thus, their conclusion should not be taken to mean that all theoretical macroeconomic implications are wrong.

We have pointed to the enormous impact of the paper of SW. During almost thirty years, their results have been quoted in papers and textbooks. One of their results is that credit rationing can only occur if the banks' return function is globally hump-shaped. Bhattacharya and Thakor (1993, p.16) emphasize this even more noting that "the key result in SW is that the bank's expected return could peak at an *interior* loan interest rate". However, it is exactly this result which is inconsistent with the very SW assumptions, as pointed out by the forthcoming publication of AR. They show that the natural outcome of the SW model is a two-price equilibrium in which only safe firms are rationed.[19] The reason is that the banks' return function *cannot* be globally hump-shaped (cf. the last paragraph in Section 2.2.2). Their paper puts the theory of equilibrium credit rationing under asymmetric information back on the research agenda.

We build on the model of SW and AR and introduce dependent project revenues as a central assumption. An implication of dependence of project revenues (and the assumption that banks pass through risk) is that households face capital risk in their consumption-savings decision. As a consequence, we have to explicitly analyze households' behavior,[20] thereby making use of the results of the theory of savings under uncertainty.

We show that there can be an equilibrium with credit rationing when project revenues are dependent.[21] In such a situation, loans are given at a single market interest rate, but some risky and some safe firms are denied credit. At that rate, safe firms have zero expected profits whereas risky firms miss a strictly positive expected profit. There is no incentive for banks

[19] See footnote 13.

[20] SW and AR assume an exogenous, increasing capital supply.

[21] AR mention two modifications of the SW assumptions which also make credit rationing à la SW possible: either a cost for seizing collateral or 'fraudulent' borrowers (cf. the AR paper for details).

to increase the interest rate on loans. This is because safe firms would not demand credit at higher rates so that there would be only risky firms left. If safe firms were out of the market, banks could extract all expected rents from active firms by increasing the interest rate on loans up to the point where risky firms have zero expected profits. As a result, the expected deposit rate that could be passed through to households would necessarily be higher than the expected deposit rate implied by the (lower) single market interest rate. However, if project revenues of the risky firms are dependent, there is more risk in the deposit rate so that such a deposit rate combination does not necessarily attract households.

As in SW, AR, and much of the other literature on equilibrium credit rationing, a 'backward-bending' capital supply is a necessary condition for rationing in our model.[22] However, there is a crucial difference. In SW, the banks' return function must be globally hump-shaped to get a backward-bending capital supply and, possibly, a credit rationing equilibrium. As mentioned above, AR have shown that such a shape is inconsistent with the SW assumptions, so that credit rationing at a single market interest rate cannot occur in the SW model. We show that a globally hump-shaped return function is not a necessary condition for a backward-bending capital supply (and, thus, for a credit rationing equilibrium) when project revenues are dependent.

We choose the simplest modelling approach with only two types of firms: risky and safe ones. We assume that only the revenues of the risky firms are dependent. In our eyes, it seems plausible to make this assumption, for the following reasons. Low-risk firms can be thought of as producers of goods which meet physiological needs. Examples for such low-risk projects are investments in industries such as foods and beverages, utilities, health care, and so on. In these industries, risk is fairly low,[23] and so is dependence. We resort to portfolio theory and the separation between market risk and idiosyncratic risk to make the argument clearer. Idiosyncratic risk is present in every firm. By definition, this kind of risk is independent

[22] In fact, capital supply has a discontinuous jump in our model since we work with only two firm types - as opposed to the continuum of types in SW and AR. Thus, capital supply is not a continuously differentiable function in our model and is not 'backward-bending' in the original sense of the expression. However, the property of a backward-bending capital supply curve which leads to equilibrium credit rationing is an 'interior' maximum, which occurs both in SW/AR and in our model.

[23] The distinction between low- and high risk industries is just one way to justify differences in risk between firms. Some descriptive statistics for global industry portfolios can be found in Table 2 in Ferreira and Gama (2005, p.203). The industries with the lowest volatilities are 'food producers & processors', 'electricity', 'investment companies', and 'food & drug retailers'. Volatilities are highest for 'IT hardware', 'software & computer services', and 'tobacco'.

between firms. The exposure to market risk, however, is highly unequal. We suggest to interpret low-risk firms as the ones which do not face significant market risk. Hence, dependence between low-risk firms is low, but idiosyncratic risk is prevalent. In contrast, due to market risk, project revenues in high-risk industries are much more likely to be highly dependent since they frequently depend on some sort of breakthrough, which might be technological, political or social in nature.[24]

The remainder of this chapter is organized as follows. In Section 2.2, we present the assumptions of our model (2.2.1) focusing on a bank's return function (2.2.2), which significantly differs from the one resulting from independent revenues. We specify the households' consumption-savings decision in a standard expected utility setup and analyze the firms' investment decision (2.2.3). In order to have a benchmark for the assessment of dependence of project revenues, we analyze possible equilibria when revenues are independent (2.2.4). The central section of the chapter describes different equilibrium cases and sets up a condition for social optimality in order to find out equilibrium inefficiencies caused by asymmetric information (2.2.5). We present comparative statics (2.2.6) before extending the model in three directions. First, in Section 2.3, we present a non-expected utility setup which allows us, amongst other things, to make the propositions from Section 2.2 more general by attributing results to different preference components. Second, Section 2.4 generalizes the concept of dependence and shows that the main results do not rely on the extreme assumption of perfectly dependent revenues among risky firms. We implement imperfect dependence as deterministic (in 2.4.1), stochastic (in 2.4.2) and, going a little further, as stochastic and uncertain (in 2.4.3).[25] As a further robustness test, we analyze a different structure of revenue dependence in Section 2.5. We add dependence of the safe firms' revenues (intra-type) and an inter-type dependence in that risky firms can only succeed if safe firms do. The final section will give some concluding remarks.

[24]An example where a technological breakthrough caused a whole industry to flourish is the IT sector. A case in point for a social breakthrough is web technology which flourished, too. Genetic engineering is an example where we do not know yet if and where a political breakthrough will occur or not.

[25]We stick to the terminology introduced by Frank Knight in 1921. In Knight (1967, pp.19-20), he writes that "...'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character...". He proposes to use "...the term 'uncertainty' to cases of the non-quantitive type". In statistical terms, this means that there is 'risk' if we have a random variable together with its distribution. If the distribution (or some parts of the distribution) is unknown, there is 'uncertainty'. The reader must be aware that other authors use the terms in a completely different way, e.g., Hubbard (2007, p.46).

2.2 The model

2.2.1 Assumptions

There is a continuum of mass N_S of safe firms and a continuum of mass N_R of risky firms. We define $\beta \equiv \frac{N_S}{N_S+N_R}$ as the share of safe firms. Project revenues $\tilde{\mathcal{R}}$ are a binary random variable: If successful, revenues of safe and risky firms are $\mathcal{R}_S$ and $\mathcal{R}_R$, respectively, where $\mathcal{R}_S < \mathcal{R}_R$.[26] In case of failure, payoffs are zero. The probability that a project succeeds is p_R for a risky and p_S for a safe firm. Note that 'safe' means relatively safe, that is, $p_R < p_S < 1$. Both types of firms have the same expected project revenue $p_R\mathcal{R}_R = p_S\mathcal{R}_S = E[\tilde{\mathcal{R}}]$. Projects require B ($< E[\tilde{\mathcal{R}}]$) units of capital which cannot be brought up internally so that firms must rely on outside funding by banks, which require C ($< B$) units of collateral. Projects are indivisible so that the only kind of credit rationing that might possibly arise is 'type II' rationing, i.e., a situation in which some borrowers receive loans, but other, identical borrowers do not.

There is asymmetric information: Firms know their type, but banks cannot observe it. However, banks know the distribution of types in the economy. Firms have one and only one project to invest in, which is either a risky or a safe one.[27] There is no moral hazard (no hidden actions). Furthermore, we assume that banks can observe revenues ex post (costless state verification), i.e., asymmetric information only exists with regard to a firm's risk type.

A central assumption concerns the dependence of project revenues: In Section 2.2, we assume that project revenues of risky firms are perfectly dependent (either all risky firms succeed or none does), whereas revenues of safe firms are independent, both among each other and w.r.t. risky firms.[28] This has consequences for pairwise correlation coefficients: Between two arbitrary risky (safe) firms, the coefficient equals one (zero). Also, pairwise correlation between any risky and any safe firm is zero.

We assume that there are many banks which take deposits from households and make

[26] There are two different types of 'returns' in our model: firms' project revenues (return on investment) and the rate of return of a bank (return on lending). Since we use the latter far more often, we reserve the more common symbol $\tilde{R}$ for it.

[27] Thus, we can speak of project type and firm type interchangeably.

[28] We relax both the assumption that only revenues of the risky are dependent and the assumption about perfect dependence, in Sections 2.4 and 2.5.

loans to firms.[29] Intermediation through banks is costless and banks have no equity.[30] Therefore, they go bankrupt (which we model as a payoff of minus infinity) if they cannot serve a stipulated claim from a deposit contract. As a consequence, banks do not take risks. Instead, they pass it through to households. They are intermediaries and, thus, active in two markets: the credit market, where they lend out funds to firms on the one hand, and the deposit market, where they collect funds from households on the other. We assume that banks set prices on the credit market, whereas they are price takers on the deposit market. The reader might wonder about the use of these two different concepts of market structure. The assumption is mainly technical in that it facilitates equilibrium analysis. Assuming price setting in both markets would require a game theoretic foundation which would tremendously increase complexity.[31] If banks are price takers, a single bank can choose an arbitrary price for its own goods (interest rate on money) without influencing any other bank's price.

The attitudes towards risk are crucial: Households are assumed to be risk-averse, whereas firms and banks are risk-neutral.[32] There are H homogeneous households whose utility is assumed to exhibit constant relative risk aversion (CRRA). They maximize expected utility in a two-period setup with exogenous income Y in period 1.[33] There is no income in period 2, only the endogenous amount of savings plus interest can be consumed. If there are several contract offers by banks, we assume that each household only invests in one contract.

Most of the assumptions are identical to the ones in a two-type version of the models of SW and AR to which we compare our results. However, the assumption of dependence of project revenues has far-reaching consequences in that it introduces capital risk for households. Thus, our second main assumption, risk aversion of households, becomes very important, too. Taken together, explicit modelling of the consumption-savings decision is indispensable.

[29] Thus, we do not analyze the raison d'être of banks. A seminal paper where bank intermediation is endogenous is Diamond (1984).

[30] If we modelled banks with equity, we would have to say much more about their risk attitude and behavior in the market. However, this shall not be our focus.

[31] As mentioned in Section 2.1, this has been done by Arnold (2007) for the case of independent project revenues and a continuum of borrower types. Results do not change.

[32] Changing risk attitudes of banks and firms does not have significant consequences for the subsequent analysis.

[33] Some readers might prefer to speak of a one-period setup since decisions are only taken at one point in time.

2.2.2 Return function

The return function of a bank is state-contingent due to dependent revenues of the risky firms.[34] Dependence is perfect so that we have two 'states (of the world)'.[35] In the good one (probability p_R), all risky firms succeed and so does a share p_S of the safe firms (if they apply for capital in the first place). With probability $(1 - p_R)$, the bad state occurs. Then, from the risky firms, banks only get the collateral, whereas of the safe firms, the same share p_S is successful (if they apply).

Expected profits of a firm j ($j \in \{S, R\}$) as a function of the loan rate r are

$$E\pi_j^{firm}(r) = (1 - p_j)(-C) + p_j \left[\mathcal{R}_j - (1 + r)B\right]. \tag{2.1}$$

Since firms are risk-neutral, they demand credit as long as their expected profits are non-negative. The respective break-even loan rates for safe and risky firms are

$$r_S = \frac{p_S \mathcal{R}_S - (1 - p_S)C}{p_S B} - 1, \tag{2.2}$$

$$r_R = \frac{p_R \mathcal{R}_R - (1 - p_R)C}{p_R B} - 1. \tag{2.3}$$

A comparison quickly shows that $r_S < r_R \Leftrightarrow E[\tilde{\mathcal{R}}] > C$, which is true since $E[\tilde{\mathcal{R}}] > B > C$ by assumption. At low interest rates ($r \in [0; r_S]$, the 'first interval'), both firm types demand credit. At high rates ($r \in (r_S; r_R]$, the 'second interval'), only the risky do, i.e., there is adverse selection. Thus, firms' expected profits can be illustrated as in Figure 2.1. It is instructive to look at the state-contingent returns of a bank in the two intervals. In the first interval, both firm types are active so that the banks' return on lending in the bad state, $i_b(r)$, is

[34] Mathematically, we should thus not speak of a 'function' but of a 'relation'. We will call it a 'function' nonetheless.

[35] By calling the two situations 'states of the world', we stick to Ingersoll (1987) and neglect Mas-Colell, Whinston, and Green (1995). The latter define a state of the world as "a complete description of a possible outcome of uncertainty, the description being sufficiently fine for any two distinct states of the world to be mutually exclusive" (p.688). Since we work with a continuum of each firm type (and of the safe type in particular), there is an infinity of states of the world differing in terms of *which* of the safe firms succeed and which fail. However, since all possible outcomes of uncertainty regarding the safe firms lead to the same prices (deposit rates), we decided to follow Ingersoll (1987, p.46) who suggests that "two or more distinguishable outcomes of nature with the same pattern of prices for the investment assets must be grouped into a single state", so that we are left with two states of the world.

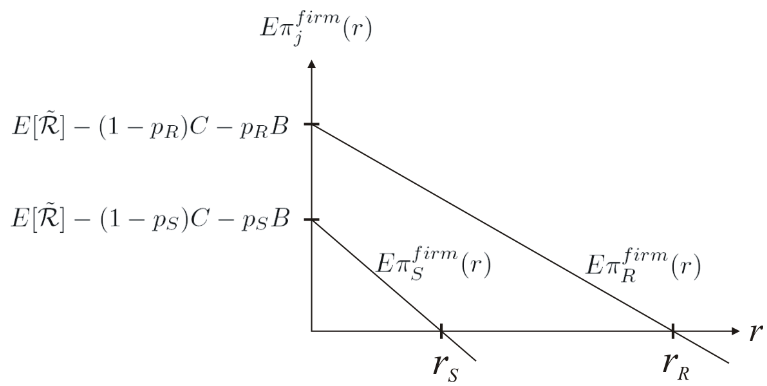


Figure 2.1: Firms' expected profits.

$$i_b(r) = \frac{N_S\left(p_S(1+r)B + (1-p_S)C\right) + N_R C}{B(N_R + N_S)} - 1,$$

and the return on lending in the good state, $i_g(r)$, is

$$i_g(r) = \frac{N_S\left(p_S(1+r)B + (1-p_S)C\right) + N_R(1+r)B}{B(N_R + N_S)} - 1.$$

These formulas are based on the assumption that the pool of borrowers in the credit portfolio of each bank is 'representative', i.e., that each bank funds the same share of risky and safe firms. From the safe firms, a share p_S is successful and pays back principal plus interest, $(1+r)B$. The remaining share $(1-p_S)$ defaults and loses its collateral. This is the same in both states of the world. By contrast, the average repayment from risky firms differs between states: $(1+r)B$ per firm in the good state and C per firm in the bad state. In the second interval, only risky firms demand capital so that

$$i_b(r) = \frac{C}{B} - 1, \qquad i_g(r) = r.$$

Expected returns of a bank in state k ($k \in \{g, b\}$, good or bad) are given by

$$E\pi^{bank}(r|k) = \frac{E[p|k](1+r)B + (1 - E[p|k])C}{B} - 1. \tag{2.4}$$

The expectation $E[p|k]$ is the expected success probability in state k and equals the proportion of successful firms, due to the law of large numbers. It is a function of the loan rate r. In Figure 2.2, the thick solid line is the expected success probability in the good state, the thick

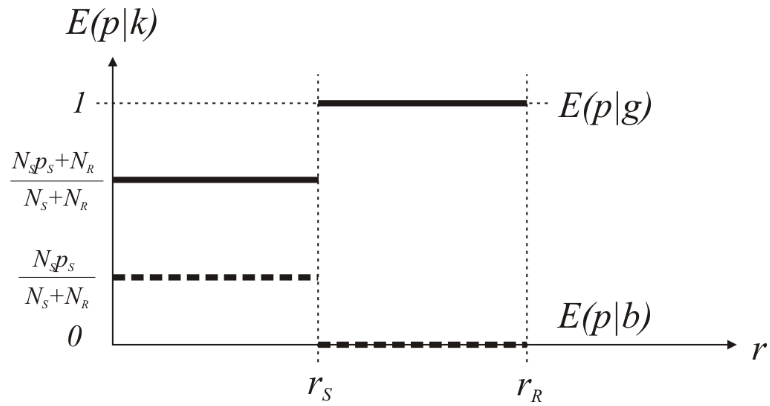


Figure 2.2: Expected success probabilities.

dashed line represents the bad state probability. $E[p|k]$ differs in (two) states and (two) intervals and can thus take on four different values.

We proceed by characterizing banks' rates of return. Several properties (of state-contingent and expected returns) can be pointed out (cf. Figure 2.3):

i) The good state return $i_g(r)$ is monotonically increasing in r with a discontinuous upward jump at r_S.

ii) The bad state return $i_b(r)$ is monotonically increasing in the first interval, but constant and at its global minimum in the second.

iii) The expected return on lending $E[i(r)]$ is monotonically increasing in r in both intervals, but jumps downwards at r_S.

iv) $E[i(r)]$ attains its global maximum at r_R.

v) The variance of the return $Var[i(r)]$ is monotonically increasing in r.

We prove these properties in Appendix 2.7.1. Property iv) is the result of AR. Intuitively, there are both risky and safe firms active at r_S, and the risky make strictly positive expected profits. At r_R, only risky firms are active and their expected profits are zero. Since safe and risky firms have the same expected revenues, expected returns of banks must be maximum at r_R.

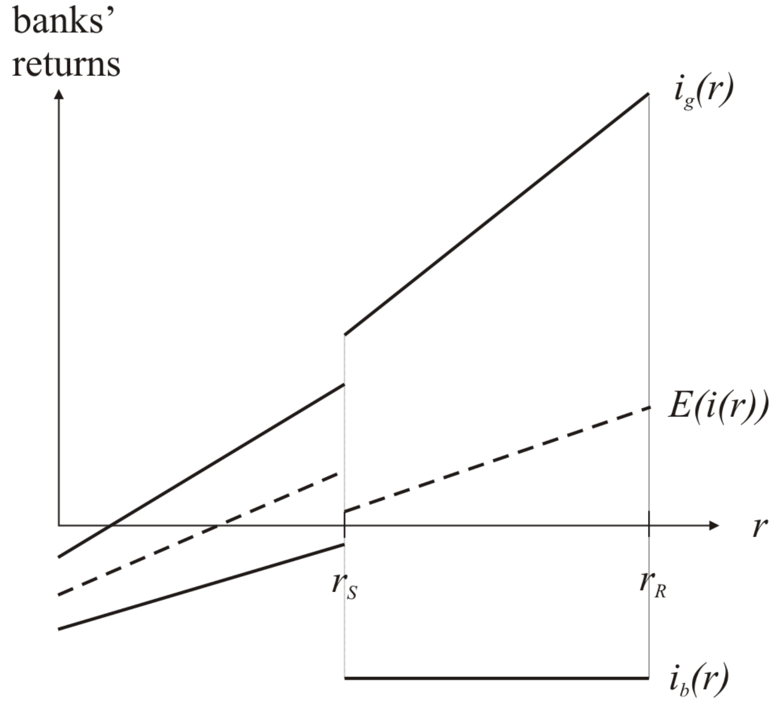


Figure 2.3: Banks' returns given dependent risky projects.

2.2.3 Credit and deposit market

From the above, it is clear that credit demand, $D(r)$, is a step function of the loan rate r, equal to $(N_S + N_R)B$ in the first and to $N_R B$ in the second interval, and zero for higher loan rates. This is because all firms have non-negative expected profits in the first interval, whereas in the second, only the risky have.

The description of the deposit market is more complicated. Since households are identical, aggregate capital supply is simply the number of households times savings of a representative household. A household's optimal amount of savings depends on the deposit rate faced. Since banks have to make zero profits in a competitive equilibrium (due to the usual downbidding process) and pass through risk, any equilibrium deposit rate combination must equal the return on lending, so that we let $i_b(r)$ and $i_g(r)$ denote both banks' rates of return and the state-contingent deposit rates offered to households. We omit the argument r and write i_g and i_b unless we talk about deposit rates at a particular loan rate, such as r_S or r_R. Households maximize expected utility. Let U denote aggregate utility of consumption over both periods, and $u(c_t)$ be instantaneous utility of consumption in period t. Using the discount factor δ

and additively-separable utility such that $U(c_1, c_2) \equiv u(c_1) + \delta u(c_2)$, optimal savings s^* solve

$$\max_s EU = E\left[u(Y - s) + \delta u(s\tilde{R})\right], \tag{2.5}$$

where $\tilde{R}$ is the random gross interest rate on deposits (not to be mixed up with $\tilde{\mathcal{R}}$). In period 1, consumption equals income minus savings. Consumption in period 2 depends on the realization of the deposit rate. The FOC is

$$u'(Y - s) = \delta E\left[u'(s\tilde{R})\tilde{R}\right]. \tag{2.6}$$

We use CRRA utility $u(c) = \frac{c^{1-\theta}}{1-\theta}$.[36] The parameter θ captures preferences both over consumption in states and in time. Since households are risk-averse, θ is positive. Optimal savings s^* can be derived from the FOC of the maximization problem,

$$s^* = \frac{Y}{1 + \left(\delta E[\tilde{R}^{1-\theta}]\right)^{-\frac{1}{\theta}}}. \tag{2.7}$$

We can replace $\tilde{R}$ using the fact that $\tilde{R} = 1+i_g$ with probability p_R and $\tilde{R} = 1+i_b$ otherwise.[37] Thus, equation (2.7) becomes

$$s^* = \frac{Y}{1 + \delta^{-\frac{1}{\theta}}\left[p_R(1+i_g)^{1-\theta} + (1-p_R)(1+i_b)^{1-\theta}\right]^{-\frac{1}{\theta}}} = \frac{Y}{1 + (\delta z)^{-\frac{1}{\theta}}}, \tag{2.8}$$

where we use the convenient definition

$$z \equiv E[\tilde{R}^{1-\theta}] = p_R(1+i_g)^{1-\theta} + (1-p_R)(1+i_b)^{1-\theta}. \tag{2.9}$$

For all $r \neq r_S$, the derivative of z w.r.t. r is

$$\frac{dz}{dr} = (1-\theta)\left[p_R(1+i_g)^{-\theta}\frac{di_g}{dr} + (1-p_R)(1+i_b)^{-\theta}\frac{di_b}{dr}\right] \gtrless 0 \Leftrightarrow \theta \lessgtr 1, \tag{2.10}$$

since $\frac{di_g}{dr} > 0$ and $\frac{di_b}{dr} \geq 0$ in each of the intervals (from properties i) and ii)). At r_S, $\frac{\Delta z}{\Delta r} \gtrless 0 \Leftrightarrow \theta \gtrless 1$. To see this, note that we have three discontinuous jumps at r_S: $1 + i_g$ goes

[36] And $u(c) = \ln(c)$ for $\theta = 1$. We will focus on $\theta < 1$ later on so that we can omit this special case.

[37] Thus, $\tilde{R}$ is a function of r, too.

up, $1 + i_b$ goes down and $E[\tilde{R}]$ goes down. Since $\tilde{R}^{1-\theta}$ is a monotonically increasing concave transformation of the binary random variable $\tilde{R}$ if $\theta < 1$, its expectation $E[\tilde{R}^{1-\theta}] = z$ must decrease at r_S. For $\theta > 1$, the transformation is monotonically decreasing and convex so that $E[\tilde{R}^{1-\theta}]$ must increase in that case.

We get indirect lifetime utility (LTU) by inserting optimal savings from equation (2.8) into the right-hand side (RHS) of equation (2.5),

$$LTU = \frac{Y^{1-\theta}}{1-\theta}\left[(\delta z)^{\frac{1}{\theta}} + 1\right]^{\theta}. \tag{2.11}$$

Aggregate savings are given by $S = Hs^*$. Both S and LTU are composite functions and can be written as $S(r) \equiv S(z[i_b(r), i_g(r)])$ and $LTU \equiv LTU(z[i_b(r), i_g(r)])$. Thus, they can be plotted in a graph with the loan rate on the abscissa. This will be important for a graphical exposition of the equilibrium. The loan rate r determines the deposit rates i_b and i_g. The latter two determine optimal savings (and, thus, capital supply) and LTU. They also determine z, which we only introduce to simplify some proofs. Regarding the difference between capital supply S and optimal savings s^*, notice that one is only an upscaled version of the other. For the equilibrium argumentation, we need capital supply, but the properties of capital supply can also be proven using the formula for optimal savings.

The equilibrium analysis crucially depends on the shape of capital supply and, thus, on the value of θ. We focus on the case of $\theta < 1$, meaning that the substitution effect outweighs the income effect in the consumption-savings decision. Thus, if $\tilde{R}$ were riskless, capital supply would be an increasing function of the deposit rate.[38] However, since $\tilde{R}$ is random, the shape (especially the slope) of capital supply depends on the change in the distribution of $\tilde{R}$.[39] A stylized graph of capital supply, LTU and the deposit rate combinations (equal to banks' state-contingent return rates) can be found in Figure 2.4. It clarifies the dependencies: Capital supply and LTU are both functions of r. There are some general properties of capital supply and LTU:

[38] When households face a higher deposit rate, two effects arise. First, since every dollar saved yields higher interest, the effect of increased savings is that households get much more future consumption in exchange for a little less current consumption. The second, opposing effect is that households can reduce savings needed for a given level of income (or consumption) in the second period.

[39] In our case, this distribution is binary. Different loan rates r yield different values for $\tilde{R}$, but the same probabilities.

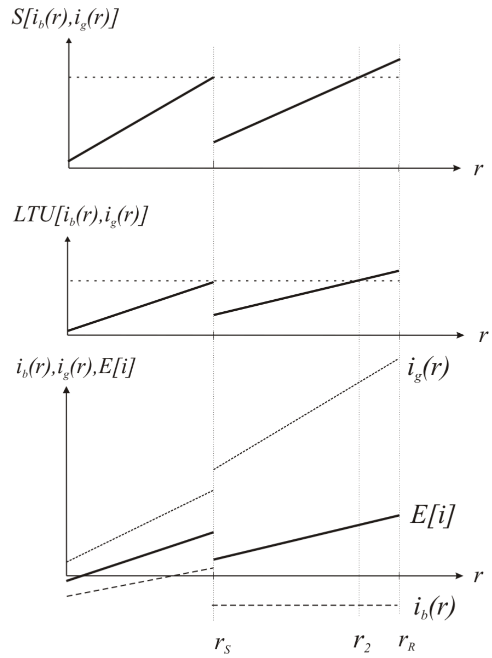


Figure 2.4: Capital supply, LTU and deposit rates.

Lemma 2.1 *If $\theta < 1$, capital supply increases monotonically in r in each of the intervals $[0, r_S]$ and $(r_S, r_R]$ and has a discontinuous downward jump at r_S.*

Proof: The fact that optimal savings s^* increase in each of the two intervals is implied by a more general result, namely Proposition 2 in Basu and Ghosh (1993, p.124). Adapting their result to the expected utility setup, their proposition states that savings are lower in a first-order stochastically dominated distribution (of deposit rates) if $\theta < 1$. From Figure 2.3 we can see that a higher loan rate within each interval implies a first-order dominant distribution. An algebraic proof starts with the fact that $\frac{dz}{dr} > 0 \Leftrightarrow \theta < 1$ from equation (2.10) in each of the two intervals. Moreover, $\frac{ds^*}{dz} > 0 \Leftrightarrow \theta < 1$ from equation (2.8). Thus, savings increase in r in each of the two intervals if $\theta < 1$.

The fact that savings decrease discontinuously at r_S follows from the observation that there must be a deposit rate combination at some loan rate $r > r_S$ which constitutes a mean preserving spread (MPS) of the deposit rate combination at r_S.[40] Rothschild and Stiglitz (1971) show that an MPS leads to a decrease in savings if $\theta < 1$. Since savings strictly

[40]If the probabilities of a binary distribution do not change, an increase in variance at a constant mean is equivalent to an MPS. For general distributions, this is not true: Any MPS implies a higher variance, but the reverse is not necessarily true.

increase within the second interval, savings must then decrease discontinuously at r_S. Algebraically, it immediately follows from $\frac{\Delta z}{\Delta r}|_{r=r_S} < 0 \Leftrightarrow \theta < 1$ and $\frac{ds^*}{dz} > 0 \Leftrightarrow \theta < 1$. q.e.d.

Lemma 2.2 *For all θ, LTU increases monotonically in r in each of the intervals $[0, r_S]$ and $(r_S, r_R]$, and has a discontinuous downward jump at r_S.*

Proof: Differentiating equation (2.11) w.r.t. z and simplifying yields

$$\frac{dLTU}{dz} = \frac{Y^{1-\theta}}{1-\theta}\left[(\delta z)^{\frac{1}{\theta}} + 1\right]^{\theta-1} \delta^{\frac{1}{\theta}} z^{\frac{1-\theta}{\theta}} \gtrless 0 \Leftrightarrow \theta \lessgtr 1, \tag{2.12}$$

since Y, θ, δ and z are always positive. From equation (2.10), we know $\frac{dz}{dr}$ in each of the two intervals: $\frac{dz}{dr} \gtrless 0 \Leftrightarrow \theta \lessgtr 1$. Therefore, $\frac{dLTU}{dr} > 0 \;\forall \theta$ in each of the two intervals.

LTU has a discontinuous downward jump at r_S since $\frac{\Delta z}{\Delta r}|_{r=r_S} \gtrless 0 \Leftrightarrow \theta \gtrless 1$ such that, for $\theta < 1$, $\frac{dLTU}{dz} > 0$ and $\frac{\Delta z}{\Delta r}|_{r=r_S} < 0$ and, for $\theta > 1$, $\frac{dLTU}{dz} < 0$ and $\frac{\Delta z}{\Delta r}|_{r=r_S} > 0$. q.e.d.[41]

Proposition 2.1 *For $\theta < 1$, capital supply and LTU have their global maximum at the same loan rate, viz., either r_S or r_R.*

Proof: From Lemmas 2.1 and 2.2, we know that both capital supply and LTU are increasing in each of the two intervals. Therefore, the global maximum of the functions is at r_S or at r_R. For $\theta < 1$, $\frac{ds^*}{dz} > 0$ and $\frac{dLTU}{dz} > 0$. Suppose that the maximum of capital supply occurs at r_S. Since $\frac{ds^*}{dz} > 0$, we must have $z(r_S) > z(r_R)$. Since $\frac{dLTU}{dz} > 0$, too, we must have $LTU(r_S) > LTU(r_R)$, i.e., the maximum of LTU occurs at r_S, too. If capital supply is maximum at r_R, $z(r_R) > z(r_S)$ and $LTU(r_R) > LTU(r_S)$, i.e., the maximum of LTU occurs at r_R, too. q.e.d.

Corollary: *If s^* is the same at two loan rates, LTU must be the same at these two rates, too.*

[41] Increasing LTU within each interval also follows from revealed preferences: Since an increase in r leads to a state-by-state dominant deposit rate combination, a change in s^* implies that households are better off. This is because households could be as well off as with the original r, simply by not changing their savings and just throwing away the additional interest on savings. Thus, the fact that households do change their savings implies that they are better off.

Proof: If s^* is the same at two loan rates, z must also be the same. Since LTU (cf. equation (2.11)) only depends on z and parameters, it must be the same at these two loan rates, too. q.e.d.

2.2.4 Independent revenues

Before defining and characterizing equilibria given dependence of project revenues in the next section, this section establishes the situation with independent project revenues as a benchmark. All other assumptions from Subsection 2.2.1 are still valid.

First, firms' decisions are unaffected by the change in the assumption. Dependence between firms' project revenues does not influence their individual success probabilities. Therefore, the break-even loan rates stay the same and adverse selection as indicated in Figure 2.1 takes place, too. Credit demand is unaffected: It is the same function as with dependent revenues, described at the beginning of Subsection 2.2.3.

On the deposit market, however, independent revenues lead to significant differences. Most importantly, the LLN implies that the return function of a bank is not state-contingent but riskless. The assumption of a continuum of both firm types implies that there is no aggregate risk and that idiosyncratic risks cancel out.

Properties iii) and iv) of the expected return function in Subsection 2.2.2 carry over to the riskless return function.[42] $E[i(r)]$ increases in each of the intervals and attains its global maximum at r_R. Households' risk aversion does not matter since they do not face capital risk.[43]

Maximizing utility is almost identical to (2.5), only using deterministic R instead of $\tilde{R}$.

$$\max_{s_{ind}} U = u(Y - s_{ind}) + \delta u(s_{ind} R), \tag{2.13}$$

where we use the index '*ind*' to indicate that the variable is derived under the assumption

[42] The proofs are identical since the formula for the riskless return function (when revenues are independent) is identical to the formula for the function of expected returns (when revenues are dependent). Starting from equation (2.4), one would have to insert unconditional expected success probabilities equal to $\frac{N_S p_S + N_R p_R}{N_S + N_R}$ in the first, and to p_R in the second interval.

[43] However, having an endogenous consumption-savings decision in an expected utility setup, we have seen that the parameter capturing risk aversion also determines households' willingness to substitute intertemporally. Thus, changes in θ do change capital supply.

that project revenues are independent. Optimal savings in equation (2.7) become

$$s^*_{ind} = \frac{Y}{1 + \delta^{-\frac{1}{\theta}} R^{\frac{\theta-1}{\theta}}}. \tag{2.14}$$

The deposit rate R is a function of the loan rate r. As in the case with dependent returns, LTU follows from inserting s^*_{ind} in the objective function on the right-hand side of equation (2.13),

$$LTU_{ind} = \frac{Y^{1-\theta}}{1-\theta} \left[\delta^{\frac{1}{\theta}} R^{\frac{1-\theta}{\theta}} + 1\right]^{\theta}. \tag{2.15}$$

As with dependent revenues, we focus on the case $\theta < 1$. The only difference in the formulas for optimal savings and LTU is that the gross deposit rate is random when firms' revenues are dependent but riskless when they are independent.[44]

Defining $z_{ind} \equiv R^{1-\theta}$, we see that z_{ind} behaves like z (both as functions of r) in that it increases in each of the intervals but decreases discontinuously at r_S for $\theta < 1$. Since s^*_{ind} and LTU_{ind} depend on z_{ind} exactly the way s^* and LTU depend on z, Lemmas 2.1 and 2.2 apply, with analogous formal proofs. However, there is an important difference in terms of Proposition 1. The fact that the return function $i(r)$ (or, from the households' point of view, the gross deposit rate R) and, thus, z_{ind} are maximal at r_R implies that savings and LTU are maximal at that rate, too.

The level of capital supply at the higher break-even loan rate r_R has another important property. In the case of independent project revenues, it can be used to determine whether it is socially desirable to fund all projects. From Subsection 2.2.2, we know that risky firms have zero expected profits at r_R so that the corresponding deposit rate $\frac{E[\tilde{\mathcal{R}}]}{B} - 1$ transfers all economic rents to households. Since safe and risky projects have the same expected revenue, the same deposit rate could be achieved from safe projects (with $r = r_S$) if there were no asymmetric information. Thus, $S(r_R)$ can be used as a measure of the socially optimal level of investment. If $S(r_R) > (N_S + N_R)B$, it is socially optimal to fund all projects.

Figure 2.5 shows possible shapes of capital demand and capital supply (the respective upper panel), and LTU (the respective lower panel). We know that capital supply and LTU are increasing in each of the intervals and that they both have a discontinuous downward

[44] Setting $E[\tilde{R}^{1-\theta}] = R^{1-\theta}$ in the RHS of equations (2.7) and (2.11) gives the RHS of equations (2.14) and (2.15), respectively.

jump at r_S. We restrict attention to cases where it is socially optimal to fund all projects, i.e., $S(r_R) > (N_S+N_R)B$. Since the maxima of both functions then occur at r_R, we distinguish between three qualitatively different shapes of both curves.

As with the variables, the index '*ind*' indicates the underlying assumption that project revenues are independent. The numbering of the three cases will become clear in the next subsection. Case III_{ind} is characterized by a low capital supply at the lower break-even loan rate r_S. Since capital supply is maximal at r_R, there must be a market-clearing loan rate $r > r_S$. Using the equilibrium concept of SW, the only equilibrium entails giving credit at the market-clearing loan rate to risky firms only (safe firms do not demand capital at that rate). In case V_{ind}, capital supply at r_S is higher than credit demand at high loan rates, but less than demand at low rates. Irrespective of there being a market-clearing loan rate (subcases a) or b)), the only possible equilibrium is a two-price equilibrium (see SW, AR, or the next subsection of this chapter). In such a situation, credit is given at two loan rates, viz., at r_S and at the loan rate $r > r_S$ at which capital supply equals capital supply at r_S (LTU is also the same at both rates). All risky firms receive credit and some safe firms are rationed. In case VI_{ind}, there is a market-clearing loan rate $r < r_S$, which is the equilibrium loan rate irrespective of there being another market-clearing loan rate. All firms get funding in that case, i.e., the social optimum arises.

2.2.5 Equilibrium

We now return to the assumption that project revenues of the risky firms are perfectly dependent so that the deposit rate is risky. As outlined above, any possible equilibrium entails zero profits for banks. Therefore, we define two different types of equilibrium with a single interest rate as follows.

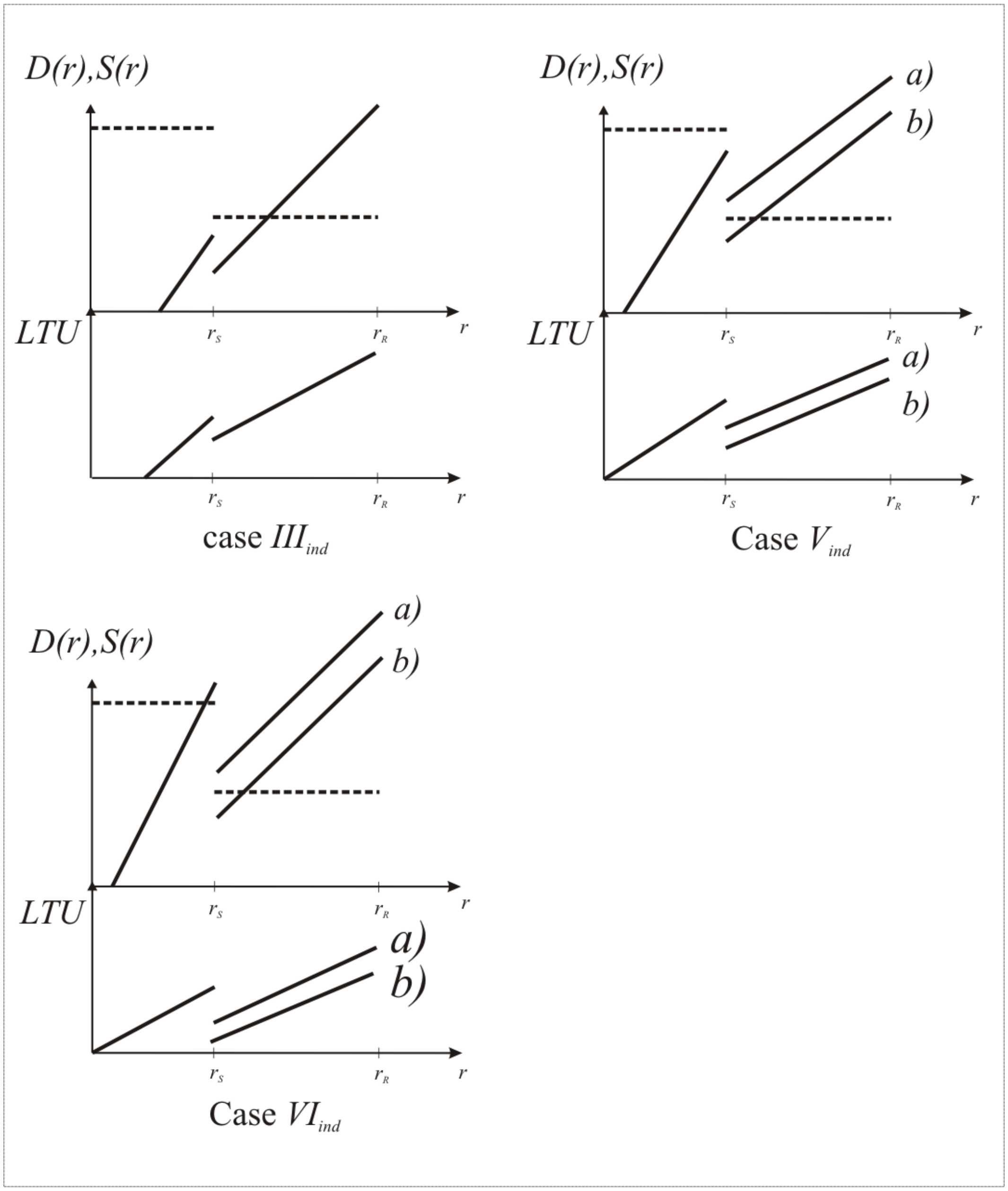


Figure 2.5: Capital supply and LTU with independent revenues.

Definition 2.1 *An 'equilibrium' is a loan rate r such that there is no r' with*

$$LTU\left[i_b(r'), i_g(r')\right] > LTU\left[i_b(r), i_g(r)\right] \tag{2.16}$$

that attracts borrowers. We distinguish between two types of equilibrium. A 'market-clearing equilibrium' has capital supply at $(i_b(r), i_g(r))$ equal to credit demand at r. A 'credit rationing equilibrium' has capital demand at r exceeding capital supply at $(i_b(r), i_g(r))$.

With regard to inequality (2.16) in Definition 2.1, there is always a profitable deviation for banks if it holds. If such a state-contingent deposit contract $(i_b(r'), i_g(r'))$ exists, households prefer that contract and a bank can thus offer a similar contract where households' utility is still higher than with the original contract, and make a profit.[45]

Due to Lemmas 2.1 and 2.2 and Proposition 2.1, we can restrict our analysis to six cases. Since households have no income in period 2 besides their own savings, CRRA utility with $\lim_{c\to 0} u'(c) = \infty$ implies that savings are positive. A graphical exposition of all six cases can be found in Figure 2.6, where the respective upper panel is a capital supply and demand diagram and the respective lower panel displays LTU, all as functions of the loan rate. Let us start with three cases in which capital supply at r_S is less than $N_R B$.

Case I:

One possibility is to have capital supply (and LTU) maximal at r_S. From Definition 2.1, it is clear that there is no market-clearing equilibrium in this case since demand exceeds supply at any loan rate r. Instead, there is a credit rationing equilibrium at the loan rate r_S. No other loan rate can be a credit rationing equilibrium since r_S yields the highest possible LTU for depositors.

Case II:

Another possibility arises if the maxima of capital supply (and LTU) are at r_R, but capital supply is still below demand at r_R. The unique equilibrium entails credit rationing. All capital that can be raised at $(i_b(r_R), i_g(r_R))$ is lent out at the loan rate r_R. It is a case of credit rationing since demand exceeds supply and there is no other loan rate with corresponding

[45] For instance, banks could offer $(i_b(r'), i_g(r') - \epsilon)$.

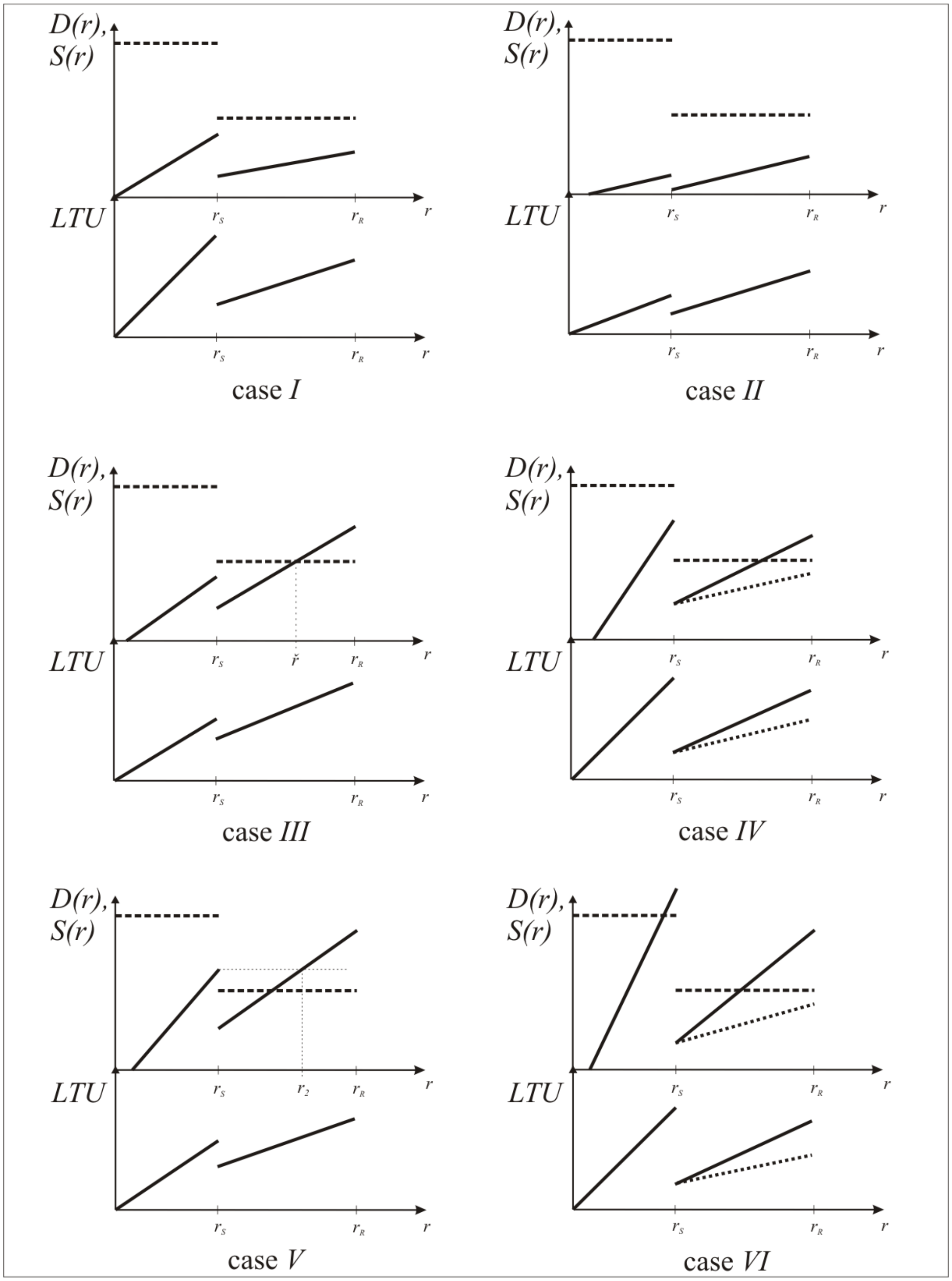


Figure 2.6: Capital supply and LTU: six equilibrium cases.

deposit rates which yield a higher LTU.

Case III:

If capital supply at r_S is less than $N_R B$, but exceeds $N_R B$ (and is thus maximal) at r_R, LTU is also maximal at r_R, and the unique equilibrium entails market clearing at a loan rate $\breve{r} \in (r_S, r_R]$ with $S(i_b(\breve{r}), i_g(\breve{r})) = D(\breve{r})$. The conditions of Definition 2.1 hold. Supply equals demand and the only loan rates $r > \breve{r}$ with corresponding deposit rates $(i_b(r), i_g(r))$ that yield a higher LTU do not attract borrowers.

Next, consider $N_R B < S(i_b(r_S), i_g(r_S)) < (N_S + N_R)B$. There are two qualitatively different subcases.

Case IV:

In the first subcase, capital supply and LTU are maximal at r_S. In this case, there is a credit rationing equilibrium at the loan rate r_S. Demand exceeds supply and there is no other loan rate with corresponding deposit rates which yields a higher LTU. It is irrelevant whether there is a higher, market-clearing loan rate or not (cf. the increasing dotted line in Figure 2.6, case IV, upper panel).

Case V:

In the second subcase, the maxima of capital supply and LTU are at r_R. The reader can check that there is neither a market-clearing nor a credit rationing equilibrium. In this case, the natural outcome of the model is a two-price equilibrium, as in SW and AR. To see this, recall from our corollary that there exists a loan rate r_2 (cf. Figure 2.4) such that $S(i_b(r_S), i_g(r_S)) = S(i_b(r_2), i_g(r_2))$ and $LTU(i_b(r_S), i_g(r_S)) = LTU(i_b(r_2), i_g(r_2))$. When households face either $(i_b(r_S), i_g(r_S))$ or $(i_b(r_2), i_g(r_2))$, their amount of optimal savings is the same and so is capital supply. Moreover, since each household can only invest in one contract, a simultaneous offer of both contracts also leads to capital supply

$S(i_b(r_S), i_g(r_S)) = S(i_b(r_2), i_g(r_2))$,[46] irrespective of how many households choose either contract. In fact, households are indifferent (ex ante) between the two deposit contracts corresponding to r_S and r_2, respectively. In analogy to SW and AR, an equilibrium situation with two loan rates can arise if banks lend out at both loan rates r_S and r_2. The amounts must satisfy three requirements. First, the sum of funds lent out at both rates (by all banks) must be equal to $S(i_b(r_S), i_g(r_S))$ (and, thus, to $S(i_b(r_2), i_g(r_2))$). Second, the amount of funds lent out (by all banks) at the higher rate r_2 must be equal to aggregate residual demand after credit has been given at the lower rate r_S. Banks lend out some funds at r_S, at which all firms apply, and subsequently lend out some funds at r_2, at which only the risky firms ask for capital. In this situation, all risky firms get capital. Some safe firms are rationed (without missing positive expected profits). Third, every single bank has to make sure that its amount of credit given at each loan rate equals its amount of deposits collected at the respective corresponding deposit rate combinations. Otherwise, a bank might not be able to meet its liabilities and, thus, incur an expected utility of minus infinity. Since households are indifferent between deposit contracts ex ante, banks are able to match loan rates with deposit rates.

Case VI:

If $S(i_b(r_S), i_g(r_S)) \geq (N_S + N_R)B$, there is a market-clearing equilibrium at or below the loan rate r_S, irrespective of the shape of capital supply and LTU at higher loan rates.

In analogy to the case of independent returns in the preceding section, we establish a condition for social desirability in order to detect any inefficiencies.[47] A well-established notion of a social optimum is the level of investment under symmetric information. The idea is that households own (the same share of all) firms so that they also know about their risk characteristics. The shortcut of using $S(r_R)$ to determine social desireability is not available here since projects differ fundamentally in the aggregate: The risky firms' project risk is a market risk and cannot be diversified.

[46] If we allowed households to invest some share of wealth in each of the contracts, this would usually not be true.

[47] The concept of social optimum defined here is just one way of modeling such an optimum. In Arnold, Reeder, and Trepl (2010), we set up a more sophisticated social optimum.

Risk-averse households prefer safe projects over risky ones since all projects have the same expected revenues. Thus, a social optimum can take on different types. It might consist of safe projects only. It can also consist of all safe and some (or all) risky projects. We use primes to label variables in a social optimum so that i is a deposit rate offered to households by banks, whereas i' is the payoff rate in a social optimum without asymmetric information. To find the socially optimal allocation, we first maximize utility assuming that households fund only safe projects.[48] If the optimal amount of the safe projects, m, is in the interval $[0, N_S]$, m is the socially optimal level of investment. However, if the solution m is larger than N_S, the optimal 'number' of safe projects is N_S.[49] Then, we find the amount of risky projects in a social optimum, $n \in [0, N_R]$, by maximizing expected utility over n, using $m = N_S$. The optimization problem in step one is

$$\max_{s'_m} U = u(Y - s'_m) + \delta u(s'_m(1 + i')), \tag{2.17}$$

where $i' = \frac{E[\tilde{\mathcal{R}}]}{B} - 1$ is the expected rate of return on a safe project, which results in a riskless payoff rate in aggregate. Using CRRA utility and solving the FOC for $(s'_m)^*$, the optimal amount of savings with only safe firms becomes

$$(s'_m)^* = \frac{Y}{1 + [\delta(1 + i')^{1-\theta}]^{-\frac{1}{\theta}}}. \tag{2.18}$$

This is sufficient to carry out

$$m = \frac{H(s'_m)^*}{B} \tag{2.19}$$

safe projects. If $m > N_S$, the optimal amount of savings with both firm types, $(s'_{mn})^*$, solves

$$\max_{s'_{mn}} EU = u(Y - s'_{mn}) + \delta \left[p_R u(s'_{mn}(1 + i'_g)) + (1 - p_R) u(s'_{mn}(1 + i'_b)) \right]. \tag{2.20}$$

[48] Note that we maximize utility as opposed to expected utility. If there are only safe projects in a social optimum, the payoff rate is riskless.

[49] A negative second derivative of the objective function w.r.t. savings is sufficient to guarantee that carrying out N_S safe projects is better than any other number of safe projects less than N_S if $m > N_S$. We have $\frac{d^2 EU}{ds^2} = u''(Y - s) + \delta E[u''(s\tilde{R}')(\tilde{R}')^2] < 0$ since $u''(c) = \frac{c^{-\theta-1}}{-\theta} < 0$ due to risk aversion. The gross payoff rate $\tilde{R}'$ is only random if the social optimum entails some risky projects.

The payoff rates depend on $n = \frac{Hs'_{mn}}{B} - N_S$. Using this, we get[50]

$$i'_g = \frac{N_S E[\tilde{\mathcal{R}}] + n\mathcal{R}_R}{(N_S + n)B} - 1 = \frac{N_S E[\tilde{\mathcal{R}}] + (\frac{Hs'_{mn}}{B} - N_S)\mathcal{R}_R}{Hs'_{mn}} - 1, \tag{2.21}$$

$$i'_b = \frac{N_S E[\tilde{\mathcal{R}}]}{(N_S + n)B} - 1 = \frac{N_S E[\tilde{\mathcal{R}}]}{Hs'_{mn}} - 1. \tag{2.22}$$

Next, using CRRA utility, we derive the FOC which we solve for $(s'_{mn})^*$ to get

$$(s'_{mn})^* = \frac{HY - \left(\frac{\delta E[\tilde{\mathcal{R}}]}{B}\right)^{-\frac{1}{\theta}} N_S(E[\tilde{\mathcal{R}}] - \mathcal{R}_R)}{\left(\frac{\delta E[\tilde{\mathcal{R}}]}{B}\right)^{-\frac{1}{\theta}} \frac{H\mathcal{R}_R}{B} + H}. \tag{2.23}$$

This uniquely determines n, i'_g and i'_b. They all depend on the number of households since an increase in the number of households results in higher savings which have to be invested in risky firms. Each household exerts an externality on all other households by making payoff rates riskier in that case. If $n > N_R$, the social optimum consists of carrying out all projects available.[51]

A crucial point of this chapter is the influence of dependence of project revenues on equilibrium outcomes. The benchmarks are the results of SW and AR, who assume independent project returns. We define a 'transition' as the change in the equilibrium case arising from the introduction of (perfectly) dependent project revenues, all other things equal, i.e., for a given parameter constellation. Note that there is a unique equilibrium case for any parameter constellation satisfying the parameter restrictions from Subsection 2.2.1: When project revenues are dependent, one of the cases from Figure 2.6 obtains. For independent revenues, Figure 2.5 shows some possible cases.

[50] We assume that banks are able to transform collateral into consumable income, whereas households are not. Thus, in contrast to the market outcome, there is no collateral C which households receive in case of default. The same is true in case a project is successful. In a two-period problem, the reader might have expected that collateral appears in the payoff rates since firms belong to households and households consume everything they possess in the second period. The assumption that only banks can transform collateral into consumable income explains why this is not the case. Instead of a 'social optimum', some readers might then prefer to speak of a 'constrained social optimum'. If we included collateral in the payoff rates, the 'insurance function' of collateral would exist in the market situation and in our social optimum so that payoff rates in each state would be unambiguously better in the social optimum.

[51] Again, the negative second derivative of the objective function w.r.t. savings is sufficient to guarantee this. This procedure is equivalent to constrained optimization with the constraint $n \leq N_R$.

Transition	Equilibrium revenues indep.	Equilibrium revenues dep.	Δ # total	Δ # safe	Δ # risky
1	III_{ind}	III	no	no	no
2	V_{ind}	V	less	less	no
3	V_{ind}	IV	less	unclear	less
4	V_{ind}	III	less	less (all to zero)	no
5	VI_{ind}	VI	no	no	no
6	VI_{ind}	V	less	less	no
7	VI_{ind}	IV	less	less	less
8	VI_{ind}	III	less	less (all to zero)	no

Table 2.1: Equilibrium transitions and amount of projects.

For now, to get clear-cut results, we focus on cases where it is socially optimal to fund all projects - both with independent and with dependent project revenues. Then, as explained in the preceding section, Figure 2.5 shows not only some but all possible equilibrium cases.

Of all cases possibly arising when revenues are dependent, we exclude cases I and II from consideration. Case I is similar to case IV: Capital supply and LTU are maximal at r_S and both are cases of credit rationing. Whether capital supply and LTU at r_S are more or less than $N_R B$ does not influence equilibrium, and neither does the existence of a market-clearing loan rate (in case IV). Moreover, case II is similar to case III: Only risky firms get credit. The difference lies in the number of risky firms in equilibrium and the loan rate charged. Thus, we are left with eight possible transitions, listed in Table 2.1.

Starting from a situation with independent revenues, the preceding analysis has shown that dependence makes the deposit rate risky without changing its expected value. Focussing on $\theta < 1$ implies that the introduction of dependent revenues leads to a decrease in savings at any loan rate r.[52]

Therefore, if parameters are such that the equilibrium with independent revenues is of case III_{ind}, the introduction of dependent revenues (all other things equal) cannot lead to a case IV, V or VI equilibrium since capital supply would have to be higher than with independent returns. Analogously, a case V_{ind} equilibrium cannot become a case VI equilibrium by the introduction of dependent project revenues.

Transition 1 is not very interesting. The allocation is similar in that only risky projects are funded. The difference is that the loan rate increases (in order to compensate households for the risk incurred). Also, transition 5 is trivial in that only the loan rate increases. Both

[52]Formally, risk leads to an MPS at any loan rate r, so that the result of Rothschild and Stiglitz (1971) applies.

allocations (with and without dependence) are socially optimal in their respective environment since only cases where the social optimum consists of funding all projects are considered. All other transitions are characterized by a change in the total number of projects funded in equilibrium. As an example, consider transition 2 in Table 2.1. If the equilibrium with both independent and dependent revenues is a two-price equilibrium, there must be fewer projects funded when revenues are dependent. This is because the total amount of funding in a two-price equilibrium equals capital supply at r_S, which is lower when revenues are dependent (column 4). Furthermore, all risky projects receive funding in every two-price equilibrium, so that there is no change in the number of risky firms (column six in Table 2.1). Therefore, the number of safe projects funded decreases (column five). The reader can go through the remaining cells of Table 2.1 to check their adequacy.

As can be seen from that table, dependence of project revenues might significantly reduce the equilibrium number of projects in a socially harmful way: While leaving unaffected the number of risky projects, (some) safe projects might not be funded. The reduction is harmful in two ways: The overall level of investment is too small and the projects being funded are of the wrong type.

Proposition 2.2 *There are parameter combinations such that a) the amount of risky projects in equilibrium is the same with dependent and independent returns, and b) the amount of safe projects in equilibrium is some strictly positive number with independent returns but less than that with dependent returns.*

Proof: Transitions 2, 4, 6 and 8 are such cases. The following examples prove that the transitions can occur.

- $p_S = 0.7, p_R = 0.5, \mathcal{R}_S = \frac{6}{0.7}, \mathcal{R}_R = 12, N_S = 100, N_R = 200, C = 0.2, B = 5, Y = 3, \delta = 0.9, H = 1200$ and $\gamma = 0.2$. Then, $D(r_S) = 1500$, $D(r_R) = 1000$, $S_{ind}(r_S) = 1266.18$, $S_{ind}(r_R) = 1981.62$, i.e., case V_{ind}. Also, $S(r_S) = 1153.97$, $S(r_R) = 1470.82$, i.e., case V. Thus, transition 2.

- $p_S = 0.8, p_R = 0.2, \mathcal{R}_S = 5, \mathcal{R}_R = 20, N_S = 25, N_R = 275, C = 0.1, B = 3, Y = 2, \delta = 0.9, H = 2000$ and $\gamma = 0.45$. Then, $D(r_S) = 900$, $D(r_R) = 825$, $S_{ind}(r_S) = 898.64$, $S_{ind}(r_R) = 2117.14$, i.e., case V_{ind}. Also, $S(r_S) = 658.20$, $S(r_R) = 1017.22$, i.e., case III. Thus, transition 4.

- $p_S = 0.8, p_R = 0.2, \mathcal{R}_S = 8, \mathcal{R}_R = 32, N_S = 25, N_R = 275, C = 0.1, B = 3, Y = 2, \delta = 0.9, H = 1700$ and $\gamma = 0.45$. Then, $D(r_S) = 900$, $D(r_R) = 825$, $S_{ind}(r_S) = 1136.93$, $S_{ind}(r_R) = 2265.74$, i.e., case VI_{ind}. Also, $S(r_S) = 848.31$, $S(r_R) = 1215.39$, i.e., case V. Thus, transition 6.

- $p_S = 0.8, p_R = 0.2, \mathcal{R}_S = 8, \mathcal{R}_R = 32, N_S = 10, N_R = 290, C = 0.1, B = 3, Y = 2, \delta = 0.9, H = 1800$ and $\gamma = 0.45$. Then, $D(r_S) = 900$, $D(r_R) = 870$, $S_{ind}(r_S) = 1088.10$, $S_{ind}(r_R) = 2399.02$, i.e., case VI_{ind}. Also, $S(r_S) = 667.35$, $S(r_R) = 1286.88$, i.e., case III. Thus, transition 8.

q.e.d.

All parameter constellations lead to social optima with all projects funded, both with independent and dependent project revenues.

The most extreme case is transition 8, where all safe projects are funded when revenues are independent but none when revenues are dependent. Intuitively, households face a trade-off: loan rate vs. firm type. A loan rate above r_S discourages safe firms from lending and makes the deposit rate riskier such that risk preferences are crucial to determine equilibrium. At the same time, such a high loan rate transfers more economic rents from risky firms to households. With few safe firms present in the economy and/or low risk aversion, the tendency is to neglect the safe firms and increase the loan rate. This is what we can observe in the parameter constellations of the proof.

Next, the transitions in Table 2.1 show that dependence of project revenues might lead to an equilibrium with credit rationing.

Proposition 2.3 *There are parameter constellations such that there is credit rationing in equilibrium with dependent revenues, but market clearing or a two-price equilibrium with independent revenues.*

Proof: Transitions 3 and 7 are such cases. They arise at several parameter specifications, as for instance:

- $p_S = 0.8, p_R = 0.2, \mathcal{R}_S = 10, \mathcal{R}_R = 40, N_S = 200, N_R = 100, C = 0.2, B = 5, Y = 2, \delta = 0.9, H = 1250$ and $\gamma = 0.2$. Then, $D(r_S) = 1500$, $D(r_R) = 500$, $S_{ind}(r_S) = 1396.62$, $S_{ind}(r_R) = 1986.64$, i.e., case V_{ind}. Also, $S(r_S) = 1386.38$, $S(r_R) = 1217.46$, i.e., case IV. Thus, transition 3.

- $p_R = 0.2, p_S = 0.99, \mathcal{R}_S = \frac{6.4}{0.99}, \mathcal{R}_R = \frac{6.4}{0.2}, N_S = 100, N_R = 200, C = 0.1, B = 5.5, Y = 2, \delta = 0.9, H = 4000$ and $\gamma = 0.37$. Then, $D(r_S) = 1650$, $D(r_R) = 1100$, $S_{ind}(r_S) = 1727.46$, $S_{ind}(r_R) = 3946.58$, i.e., case VI_{ind}. Also, $S(r_S) = 1625.20$, $S(r_R) = 1602.80$, i.e., case IV. Thus, transition 7.

q.e.d.

Transition 3 is the case where a two-price equilibrium becomes an equilibrium with credit rationing, transition 7 has a market-clearing equilibrium turned into a credit rationing equilibrium. This is particularly interesting in light of the recent result of AR: Credit rationing as in cases I and IV is impossible in the SW setup.[53] Proposition 2.3 shows that the intro-

[53] Recall from the discussion of case V that a two-price equilibrium has rationing, too. However, only safe firms are rationed, whereas in a credit rationing equilibrium (at a single interest rate), both firm types are rationed, and the risky miss a strictly positive expected profit.

duction of dependence of project revenues might make credit rationing possible again. This happens in spite of it being socially optimal to fund all projects.

Up to now, we focussed on cases in which the social optimum consists of all projects. We then compared the situation with and without dependence of project revenues. Next, we look at cases where the social optimum does not necessarily consist of all projects.[54]

There are two kinds of inefficiencies: the number of projects in equilibrium and the type of project.[55]

Proposition 2.4 *With dependent project revenues, the number of safe projects in equilibrium cannot be higher than in a social optimum.*

Proof: We prove this for each equilibrium case separately. For cases II and III, the assertion is true since the equilibrium in both cases consists of only risky projects and the social optimum always has some safe projects funded.

For cases I, IV and VI, we only have to look at social optima with $m < N_S$ safe projects since there cannot be more than N_S safe firms funded in any equilibrium. We show that $m < N_S$ implies that the total number of projects in equilibrium and, thus, the number of safe projects in particular, is less than the number of safe projects in the social optimum.

To see this, note that the payoff rate in a social optimum is riskless (and equal to $\frac{E[\tilde{\mathcal{R}}]}{B} - 1$), whereas the deposit rate in equilibrium is always risky with a mean below the payoff rate (from Figure 2.3, recall that $\frac{E[\tilde{\mathcal{R}}]}{B} - 1$ is the highest possible deposit rate mean and it only occurs in equilibrium case II, i.e., $E[i(r)] < \frac{E[\tilde{\mathcal{R}}]}{B} - 1$ for $r \leq r_S$). Since we focus on $\theta < 1$, the substitution effect outweighs the income effect. Therefore, if the equilibrium expected deposit rate were riskless, savings in equilibrium would be lower than in a social optimum. The additional effect of risk reduces equilibrium savings even further, a result well known from Rothschild and Stiglitz (1971).[56]

[54]This can be considered the more natural case since parameter constellations which lead to a social optimum with funding for all projects in our six cases are rather extreme, especially collateral C is very small relative to capital requirement B.

[55]For all equilibrium cases, we give some numerical examples together with the respective amounts of safe and risky projects both in equilibrium and in the social optimum in Appendix 2.7.2.

[56]Therefore, instead of repeating this chain of arguments each time we compare two different deposit (or payoff) rate combinations, we can use mean-variance argumentation as a shortcut in our setup with a binary distribution of the deposit (or payoff) rate and CRRA utility.

The same logic is true for case V, the two-price equilibrium: Having a social optimum with $m < N_S$, the deposit rate combination at either equilibrium loan rate has a mean which is lower than the riskless payoff rate in optimum (i.e., $E[i(r)] < \frac{E[\tilde{\mathcal{R}}]}{B} - 1$), and a strictly positive variance. Thus, total savings in equilibrium and, in particular, the number of safe projects must be lower. q.e.d.

However, it is interesting to note that this does not imply that the total number of projects in equilibrium is always lower than in a social optimum. There might be overinvestment.

Proposition 2.5 *The equilibrium level of investment in each of the six cases might be higher than the respective socially optimal level.*

Proof: As can be seen from Table 2.4 in Appendix 2.7.2, there are parameter constellations which yield a social optimum with a lower level of investment than in the respective equilibrium for each of the six cases. q.e.d.

If the social optimum consists of $m < N_S$ firms, this cannot happen (cf. the argument from the proof of Proposition 2.4). If $n > 0$, the mean of the payoff rates in the social optimum is always (equal to $\frac{E[\tilde{\mathcal{R}}]}{B} - 1$ and thus) higher than (or, in case II, equal to) the mean of any equilibrium deposit rate combination. The variance, however, can also be higher.

The most important determinant of the variance is the composition of the respective loan portfolio. The more safe firms relative to risky firms, the lower the variance. In equilibrium, this relationship is $\frac{N_S}{N_R}$ in cases I, IV and VI, less than that in case V, and zero in cases II and III. In the social optimum, this ratio is $\frac{m}{n}$. If the social optimum consists of only safe firms, this ratio approaches infinity and is thus higher than in any equilibrium. Even if the social optimum consists of all risky and all safe firms, this ratio equals $\frac{N_S}{N_R}$. From this point of view, the variance of payoff rates should never be higher than the variance of deposit rates.

However, there is another important factor, viz., the amount of collateral. A consequence of our definition of a social optimum is that an increase in collateral leaves the variance (and the mean anyway) of the payoff rates unaffected but decreases the variance of the deposit rate combinations. In the market setup, collateral has an insurance function. Therefore, for parameter constellations with high C and/or high ratio $\frac{N_S}{N_R}$, the variance of the deposit rate can be lower than the variance of the payoff rates.[57]

In sum, the effect of the higher risk in a social optimum can then outweigh the effect of the higher mean such that, with $\theta < 1$, savings in equilibrium can be higher than in a social optimum.

Another natural question arises: Can the equilibrium allocation be the one of a social optimum so that number and type of projects coincide, i.e., is there an efficient equilibrium?

Proposition 2.6 *Market clearing at a loan rate $r \leq r_S$ is a necessary (but not sufficient) condition for a socially efficient equilibrium.*

Proof: In case VI, all firms get funding in equilibrium. If the social optimum consists of all projects, the equilibrium allocation is efficient. Otherwise, it is not. The former occurs, for instance, when parameters are as in line 11 in Table 2.3 in Appendix 2.7.2, the latter occurs, e.g., for parameters as in line 12 (the allocations can be found in the respective lines in Table 2.4 in Appendix 2.7.2.).

Evidently, cases II and III cannot be socially efficient since only risky firms are funded. The equilibrium in cases I, IV and V cannot be efficient if $m = N_S$ since there is always rationing of safe firms. If $m < N_S$, the arguments in the proof of Proposition 2.4 can be used to see that the number of safe firms in equilibrium is lower than the number of safe firms in a social optimum. q.e.d.

We conclude this subsection with a short summary. First, we defined equilibrium and described possible cases. Next, the concept of social desirability was developed for the case of dependent project revenues. Assuming that it is socially optimal to fund all projects, we found that dependence of project revenues can have two significant effects: First, it might reduce the number of safe firms in equilibrium and, second, it can lead to credit rationing in equilibrium. We stressed the importance of a closer examination of number and type of projects in a social optimum. The interplay between asymmetric information and dependent project revenues can have serious consequences: The number of safe firms in equilibrium is never higher than the socially optimal amount. However, it is possible to have an equilibrium with

[57]The reader is invited to check this statement with the parameter constellations in Table 2.4 in Appendix 2.7.2.

more projects than in a social optimum. The only efficient equilibrium is a market-clearing one with all firms active.

2.2.6 Comparative statics

Two important assumptions underlying our model are dependent revenues and households' risk aversion. We have analyzed the influence of perfectly dependent revenues for given degrees of risk aversion (and intertemporal substitution preferences at the same time). The aim of this subsection is to see how changes in θ influence the equilibrium, given perfect dependence of project revenues (of the risky firms). An increase in θ has two effects: higher risk aversion and stronger preferences for intertemporally smooth consumption.[58] As a consequence, changes in capital supply and LTU might cause changes in equilibrium. In Figure 2.6, we have seen that the question of whether the maximum of capital supply (and LTU) occurs at r_S or r_R is of utmost importance for the equilibrium. If $S(r_R) - S(r_S)$ is positive ($\Leftrightarrow s^*(r_R) - s^*(r_S) > 0$), the maxima occur at r_R, and vice versa.

At $\theta = 0$, the difference $s^*(r_R) - s^*(r_S)$ is either 0 or Y (see Figure 2.7). $\theta = 0$ means risk neutrality so that only the expected deposit rate matters for the consumption-savings decision. Furthermore, the marginal utility of consumption is finite and equal to 1 for all consumption levels, in particular $u'(0) = 1$ in both periods. Therefore, the FOC (2.6) for $\theta = 0$ becomes $1 = \delta(1 + E[i])$. There is a critical $\overline{E[i]}$ from which on (discounted) marginal utility in period two will be higher than in period 1 so that households will save Y for $E[i] > \overline{E[i]}$ and nothing otherwise. Since $E[i(r_S)] < E[i(r_R)]$ (property iv) in Subsection 2.2.2), $s^*(r_R) - s^*(r_S)$ starts at either 0 or Y.[59] The former occurs if $\overline{E[i]} < E[i(r_S)]$ or $\overline{E[i]} > E[i(r_R)]$, the latter if $E[i(r_S)] < \overline{E[i]} < E[i(r_R)]$. For different parameter constellations, we get four qualitatively

[58] θ is the Arrow-Pratt measure of relative risk aversion and the inverse of the elasticity of intertemporal substitution.

[59] In a non-generic case, namely $\overline{E[i]} = E[i(r_S)]$, the difference at $\theta = 0$ can be anything between 0 and Y.

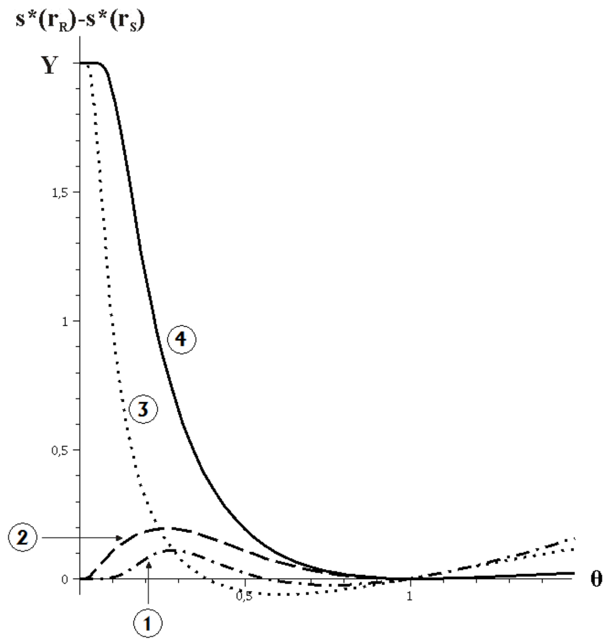


Figure 2.7: Four different shapes of $s^*(r_R) - s^*(r_S)$.

different shapes for $s^*(r_R) - s^*(r_S)$ (as a function of θ) which we plot in Figure 2.7.

1. The curve starts at 0, increases up to a local maximum, decreases to a negative local minimum and becomes zero at $\theta = 1$.

2. It starts at 0, increases up to a local maximum, decreases to a root at $\theta = 1$ so that there is no intersection with the abscissa for $\theta < 1$.

3. It starts at Y, decreases to a (negative) local minimum and becomes zero at $\theta = 1$.

4. It starts at Y and decreases to a root at $\theta = 1$ so that there is no intersection with the abscissa for $\theta < 1$.

The root at $\theta = 1$ always occurs since $\theta = 1$ means log utility with the well-known property of a constant amount of savings (cf. footnote 36). With most parameter constellations, we get shape 1 or 3, i.e., there is a critical $\theta < 1$ above which the maximum of capital supply (and LTU) occurs at r_S. However, shapes 2 and 4 show that such a critical θ does not necessarily exist.[60] If it were possible to vary risk aversion alone, an educated guess would be to always

[60] An exemplary parameter constellation which leads to shape 2 is: $p_S = 0.9, p_R = 0.1, \mathcal{R}_S = 10, \mathcal{R}_R = 90, N_S = 200, N_R = 280, C = 2, B = 5, Y = 2, \delta = 0.5$ and arbitrary H. For shape 4: $p_S = 0.9, p_R = 0.1, \mathcal{R}_S = 10, \mathcal{R}_R = 90, N_S = 200, N_R = 280, C = 2, B = 5, Y = 2, \delta = 0.9$ and arbitrary H.

have a critical value from which on the high risk at r_R leads to $S(r_R) < S(r_S)$. This is the most important drawback of the preceding analysis: We cannot vary risk aversion keeping all other things equal. A change in risk aversion also changes time preferences since both are governed by the same parameter, θ. Therefore, we present a non-expected utility setup (non-EU) in the next section.

2.3 Extension I: non-expected utility

Fortunately, there is a way to separate risk aversion and intertemporal preferences. Selden (1978, 1979) implemented the two-period consumption-savings decision with two parameters, each measuring one of the two preference components.[61]

2.3.1 Capital supply, lifetime utility and equilibrium

Selden proposes to use the certainty equivalent of uncertain[62] consumption in the second period to maximize overall utility. He calls his approach the ordinary certainty equivalent representation of preferences. Agents solve

$$\max_s U = \{u(Y - s) + \delta u(\hat{c}_2)\}, \tag{2.24}$$

where $\hat{c}_2$ is the certainty equivalent of consumption defined by $v(\hat{c}_2) = E[v(s\tilde{R})]$, so that $\hat{c}_2 = v^{-1}(E[v(s\tilde{R})])$. As before, $s\tilde{R}$ is the uncertain income (and consumption) flow in the second period. The function $v(\cdot)$ is another utility function introduced to determine the certainty equivalent of random consumption in the second period and is assumed to be of the CRRA type: $v(c) = \frac{c^{1-\gamma}}{1-\gamma}$. Due to households' risk aversion, $\gamma > 0$. Note that we impose no upper bound on γ, in particular, it might exceed one. Instantaneous utility in the respective period is $u(c) = \frac{c^{1-\alpha}}{1-\alpha}$, which implies constant elasticity of substitution (CES) time preferences.[63] The optimization problem as a whole is a non-expected utility approach since

[61] Among others, Selden's work prompted the often-quoted paper by Epstein and Zin (1989), which might be more familiar to readers. However, Epstein and Zin work with an infinite horizon, which does not fit in the model structure chosen in this chapter.

[62] As will become clear later on in this Subsection, households are assumed to know the distribution of the deposit rate so that, sticking to Knight's distinction between risk and uncertainty, we should rather speak of 'risky consumption', see footnote 25.

[63] The same mathematical function displays both CRRA and CES. We omit the special cases of $u(c) = \ln c$ for $\alpha = 1$ since we restrict our attention to $\alpha < 1$ later on, and do not look at $v(c) = \ln c$ for $\gamma = 1$ either,

the objective function is in general not linear in probabilities.[64] Households have a desire for smooth consumption so that $\alpha > 0$. High γ indicates high risk aversion, whereas high α indicates a low intertemporal elasticity of substitution, i.e., there is a strong desire for a smooth consumption path. Using the above utility functionals and the definition of $\hat{c}_2$, we get the FOC of equation (2.24),

$$(Y-s)^{-\alpha} = \delta E[(s\tilde{R})^{-\gamma}]E[\tilde{R}]\left[s(E[\tilde{R}^{1-\gamma}])^{\frac{1}{1-\gamma}}\right]^{\gamma-\alpha}. \tag{2.25}$$

Solving for optimal savings s^* yields

$$s^* = \frac{Y}{1+\delta^{-\frac{1}{\alpha}}\hat{R}^{\frac{\alpha-1}{\alpha}}}, \tag{2.26}$$

where

$$\hat{R} = \left(E[\tilde{R}^{1-\gamma}]\right)^{\frac{1}{1-\gamma}} \tag{2.27}$$

is the certainty equivalent interest rate. This is the riskless interest rate which makes an agent with CRRA as well off as the uncertain payoff $\tilde{R}$. Plugging s^* into the RHS of equation (2.24) and using $\hat{c}_2 = \hat{R}s^*$, simplification of the resulting expression yields LTU as a function of $\hat{R}$,

$$LTU = \frac{Y^{1-\alpha}}{1-\alpha}\left(\delta^{\frac{1}{\alpha}}\hat{R}^{\frac{1-\alpha}{\alpha}}+1\right)^{\alpha}. \tag{2.28}$$

Lemma 2.3 *Using non-expected utility, LTU and, if $\alpha < 1$, capital supply increase monotonically in r in each of the intervals $[0, r_S]$ and $(r_S, r_R]$ with a discontinuous downward jump at r_S. If $\alpha < 1$, capital supply and LTU have their global maximum at the same loan rate, viz. either r_S or r_R. If s^* is the same at two loan rates, LTU must be the same at these two rates, too.*

In Appendix 2.7.3, we prove all these properties. The assertions in Lemma 2.3 correspond to Lemma 2.1, Lemma 2.2, Proposition 2.1 and its corollary in the expected utility setup in Subsection 2.2.3. Thus, we can see that the parameter for intertemporal substitution (α) alone

except if necessary in mathematical proofs.

[64] Setting $\gamma = \alpha = \theta$, we are back in the expected utility case as in Section 2.2. Note that the definition of the certainty equivalent follows pure expected utility theory.

is responsible for the slope of capital supply. As long as $\alpha < 1$, capital supply is increasing in r in each of the two intervals with a downward jump at r_S, irrespective of γ, the parameter capturing risk aversion. However, at the same time,

Lemma 2.4 *Whether the maximum of capital supply and LTU occurs at r_S or r_R does not depend on the intertemporal substitution parameter α.*

The proof of this lemma is also delegated to Appendix 2.7.3. As in the expected utility setup in the previous section, we focus on an increasing capital supply. Thus, we only look at $\alpha < 1$. As a consequence, we get the same six equilibrium cases as in the expected utility setup, depicted in Figure 2.6.

2.3.2 Comparative statics

Proposition 2.7 *If $\alpha < 1$, an increase in households' risk aversion decreases optimal savings at any combination of i_b and i_g.*

Proof: First, from Basu and Ghosh (1993), we know that a decrease in $\hat{R}$ implies a reduction in savings if $\alpha < 1$. This can be seen from differentiating optimal savings in (2.26) w.r.t. $\hat{R}$,

$$\frac{ds^*}{d\hat{R}} = Y \left[\frac{\delta^{\frac{1}{\alpha}} \hat{R}^{\frac{1-2\alpha}{\alpha}}}{\left(1 + \delta^{\frac{1}{\alpha}} \hat{R}^{\frac{1-\alpha}{\alpha}}\right)^2} \frac{1-\alpha}{\alpha} \right]. \tag{2.29}$$

Therefore, it is sufficient to show that $\frac{d\hat{R}}{d\gamma} < 0$ for $\alpha < 1$. Since the logarithm is a monotonic transformation, this is equivalent to showing that $\frac{d \ln \hat{R}}{d\gamma} < 0$. From the definition of $\hat{R}$ in equation (2.27), it follows that

$$\ln \hat{R} = \frac{1}{1-\gamma} \ln E[\tilde{R}^{1-\gamma}].$$

Using the product rule and applying logarithmic differentiation, the derivative becomes

$$\frac{d \ln \hat{R}}{d\gamma} = \frac{\ln E[\tilde{R}^{1-\gamma}]}{(1-\gamma)^2} - \frac{1}{1-\gamma} \frac{E[\tilde{R}^{1-\gamma} \ln \tilde{R}]}{E[\tilde{R}^{1-\gamma}]}.$$

Defining $\phi \equiv \tilde{R}^{1-\gamma}$, this equation can be written as

$$\frac{d\ln\hat{R}}{d\gamma} = \frac{\ln E[\phi]}{(1-\gamma)^2} - \frac{1}{1-\gamma}\frac{E[\phi\frac{1}{1-\gamma}\ln\phi]}{E[\phi]} = \frac{1}{(1-\gamma)^2 E[\phi]}\left\{E[\phi]\ln(E[\phi]) - E[\phi\ln(\phi)]\right\}.$$

The difference in braces on the far RHS of the above equation is decisive since the factor in front of it is positive. Defining a new function $f(\phi) \equiv \phi\ln(\phi)$, the term in braces is negative if f is strictly convex in ϕ. This is because a strictly convex function f satisfies $E[f(\phi)] > f(E[\phi])$. Since $f'(\phi) = \ln(\phi) + 1$ and $f''(\phi) = \phi^{-1} > 0$, f is convex. q.e.d.

This proof holds for arbitrary distributions of $\tilde{R}$. So if risk aversion increases, optimal savings decrease at any (combination of deposit rates corresponding to a) loan rate r. We can also say something about the location of the maxima of capital supply and LTU, which is crucial to determine the equilibrium.

Proposition 2.8 *Capital supply and LTU attain their respective maxima at r_S for sufficiently high risk aversion. For risk aversion sufficiently low, capital supply and LTU attain their respective maxima at r_R.*

Proof: Let $f(\gamma) \equiv s^*(\gamma)|_{r=r_R} - s^*(\gamma)|_{r=r_S}$. From AR, we know that the expected return function attains its global maximum at r_R. This implies that, given $\alpha < 1$, capital supply must be maximum at r_R for risk-neutral households ($\gamma = 0$) so that $f(0) > 0$. The other extreme is an infinitely high aversion to risk. In this case, we can see that savings are higher at r_S by looking at the limit of the difference in savings,

$$\lim_{\gamma\to\infty} f(\gamma) = \lim_{\gamma\to\infty}\left[\frac{Y}{1+\delta^{-1/\alpha}(\hat{R}(r_R))^{\frac{\alpha-1}{\alpha}}} - \frac{Y}{1+\delta^{-1/\alpha}(\hat{R}(r_S))^{\frac{\alpha-1}{\alpha}}}\right],$$

where $\hat{R}(r_j)$ is the certainty equivalent interest rate for the state-contingent deposit rates $i_g(r_j)$ and $i_b(r_j)$ ($j \in \{S, R\}$) with their corresponding probabilities p_R and $(1-p_R)$. Since γ appears only in $\hat{R}$, we only have to look at the limits of $\hat{R}$ at r_R and r_S. $\hat{R}$ decreases in γ (cf. the proof of Proposition 2.7), and is applied to a Bernoulli lottery. For higher risk aversion, the certainty equivalent interest rate approaches the worse of the two lottery outcomes. In the limit, $\hat{R}(r_j) = (1 - i_b(r_j))$, i.e., in order to avoid risk, a maximally risk-averse household

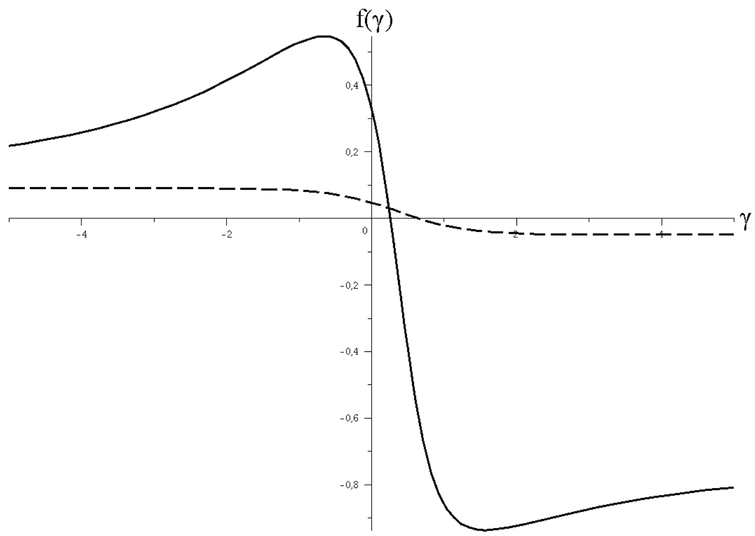


Figure 2.8: Two examples of shapes of $f(\gamma)$.

accepts the worse of the two lottery outcomes.[65] Since $i_b(r_R) < i_b(r_S)$ (cf. property ii) in Subsection 2.2.2), we get

$$(1 + i_b(r_R))^{\frac{\alpha-1}{\alpha}} > (1 + i_b(r_S))^{\frac{\alpha-1}{\alpha}},$$

given that $\alpha < 1$. Therefore, the above limit is negative, and its algebraic form can be obtained by substituting $(1 + i_b(r_i))$ for $\hat{R}(r_i)$. q.e.d.

Thus, $f(\gamma)$ starts at some positive value for $\gamma = 0$ and asymptotically approaches the negative limit for $\gamma \to \infty$, either from above or from below. Different possible shapes of $f(\gamma)$ can be seen in Figure 2.8. Note that we show two examples without being exhaustive.[66] Since $f(\gamma)$ is a continuous function, there is at least one root. From the graphs in Figure 2.8,[67] it seems that there is only one critical value of γ from which on the maximum of capital supply occurs at r_S. For a rigorous proof, we would have to preclude the possibility of more than one root, which is not straightforward (and which we were not able to achieve).

We illustrate Proposition 2.8 in Figure 2.9. Capital supply and LTU are plotted for three

[65] The opposite is also true: For maximum risk lovers, i.e., $\gamma \to -\infty$, the better of the two outcomes, $(1 + i_g(r_j))$, is the certainty equivalent interest rate.

[66] For extreme parameter constellations, we found cases where $f'(0)$ is positive, i.e., the maximum of the difference in savings might occur at a positive value of γ.

[67] Apart, we tried a large amount of numerical examples. In none of these, there was more than one root.

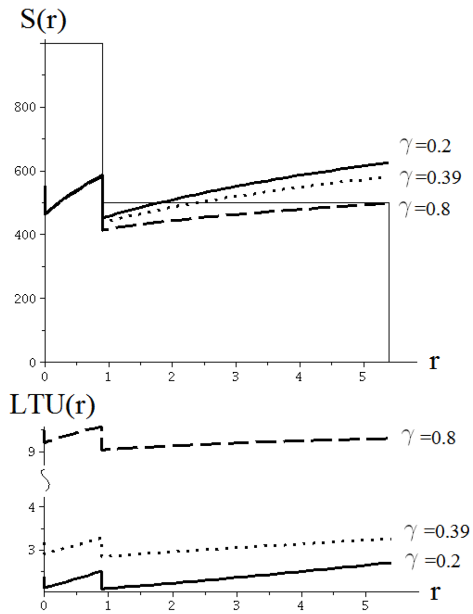


Figure 2.9: Capital supply and LTU for different values of γ.

different values of γ, as indicated in that figure. We keep all other parameters constant.[68] The upper panel depicts capital supply and demand, the lower panel shows LTU, both depending on the loan rate r. For a high value of γ, both capital supply and LTU have their maximum at r_S (the dashed curves). For low γ, both maxima occur at r_R (the solid curves). At $\gamma \approx 0.39$, capital supply (and LTU) are the same at r_S and r_R (the dotted curves).

Proposition 2.8 and the graphs in Figure 2.8 support the guess we were making in Subsection 2.2.6: There is a critical value for the risk-aversion parameter from which on the maximum occurs at r_S, so that the type of equilibrium might change. Apart from this, the non-expected utility setup has further convenient features: On the one hand, it is a genuine generalization since setting $\alpha = \gamma = \theta$ yields the expected utility setup. This added flexibility allows to vary two preference components separately. In particular, we are able to vary the degree of risk aversion to arbitrarily high values without changing the sign of the slope of capital supply. On the other hand, the empirical literature shows that there is no unanimous relationship between risk aversion and the intertemporal elasticity of substitution. In particular, the hypothesis of an inverse relationship as implied by the expected utility setup

[68] We use $p_S = 0.8$, $p_R = 0.2$, $\mathcal{R}_S = 10$, $\mathcal{R}_R = 40$, $N_S = 100$, $N_R = 100$, $C = 2$, $B = 5$, $Y = 2$, $H = 500$, $\alpha = 0.5$ and $\delta = 0.9$.

is rejected. However, there is dissent regarding the plausible magnitudes of α and γ. In an empirical study, Attanasio and Weber (1989) get $\alpha < 1$ and $\gamma > 1$, which could not be considered using an expected utility setup.[69]

2.4 Extension II: imperfect dependence

When we were talking about dependent project revenues in the above analysis, we considered perfect dependence: Either all risky firms succeed or all fail. This is the most extreme sort of dependence. In this section we introduce a random variable $\tilde{q}$ with support $[0, 1]$ to allow for variations in the degree of dependence. Let $f(\tilde{q})$ be its density. We can interpret $\tilde{q}$ as an aggregate shock which determines capital risk for households.[70] The extreme realization $q = 0$ means that all risky (and, thus, all) firms' revenues are independent, such that we get the SW setup with a riskless deposit rate. The other extreme realization $q = 1$ means that all risky firms have perfectly dependent revenues (as in the previous section). Intermediate values of q yield imperfect dependence among the risky firms. For example, if $q = 0.5$, half the risky firms have independent revenues, but the other half will either succeed or fail altogether. Throughout this chapter, the revenues of the safe firms are independent so that a share p_S of them succeeds. The timing of this new model structure is illustrated in Figure 2.10.

The analysis in this section is based on expected utility maximization of households as in Section 2.2.

2.4.1 Deterministic degree of dependence

In this subsection, we assume that households take the consumption-savings decision after $\tilde{q}$ has been realized. Evidently, credit demand of firms is unaffected. The realization of $\tilde{q}$ only changes aggregate risk and, thus, the deposit rates. Banks act as before, i.e., they pass through risk and make zero profits in any potential equilibrium. Thus, the good and bad state deposit rates, which still occur with probabilities p_R and $(1 - p_R)$, become

[69] We do not want to omit contrary studies. Hall (1987) and Epstein and Zin (1991) estimate values of $\alpha > 1$. If $\alpha > 1$, the whole SW and AR analysis which is based on an increasing capital supply breaks down.

[70] The project revenue is another sort of random variable which can be interpreted as a shock to each individual firm.

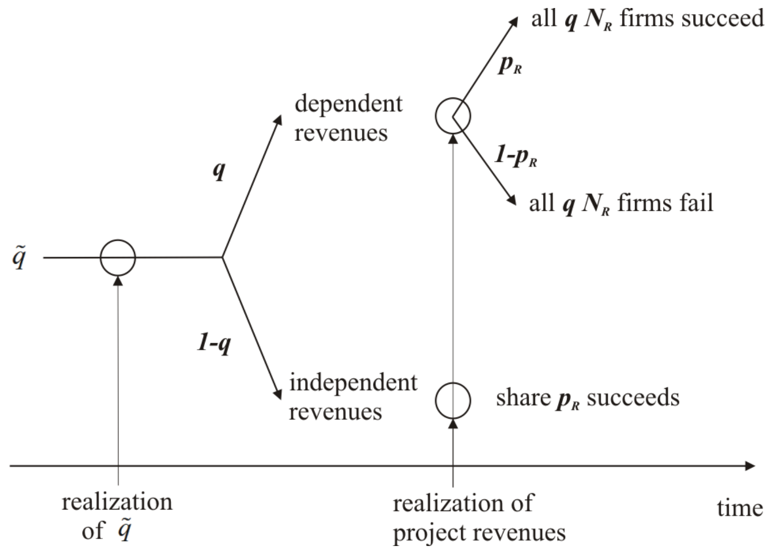


Figure 2.10: Timing of the new model structure.

$$i_b(q) = (1-q)E[i] + qi_b, \qquad\qquad i_g(q) = (1-q)E[i] + qi_g. \tag{2.30}$$

As in Subsection 2.2.3, i_b and i_g are the deposit rates for perfectly dependent revenues and $E[i]$ is the expected deposit rate. Also, i_b, i_g, and $E[i]$ depend on r which we again omit, unless stated otherwise. The optimal amount of savings resulting from expected utility maximization becomes a function of q.

$$s^*(z(q)) = \frac{Y}{1 + (\delta z(q))^{-\frac{1}{\theta}}}, \tag{2.31}$$

where $z(q) \equiv p_R(1 + (1-q)E[i] + qi_g)^{1-\theta} + (1-p_R)(1 + (1-q)E[i] + qi_b)^{1-\theta}$, in analogy to the definition of z in Subsection 2.2.3. Instead of being a function of z, optimal savings are a function of $z(q)$. The same is true for indirect lifetime utility: Take equation (2.11) and write $z(q)$ instead of z to get $LTU(z(q))$.

For the equilibrium analysis, we need to know how capital supply and LTU behave as functions of r. They depend on the shape of the deposit rate combinations as functions of r (cf. Figure 2.3). We can determine capital supply and LTU looking at the influence of a change in q at a given r. A decrease in q does not change the expectation of the deposit rate, but decreases its variance.

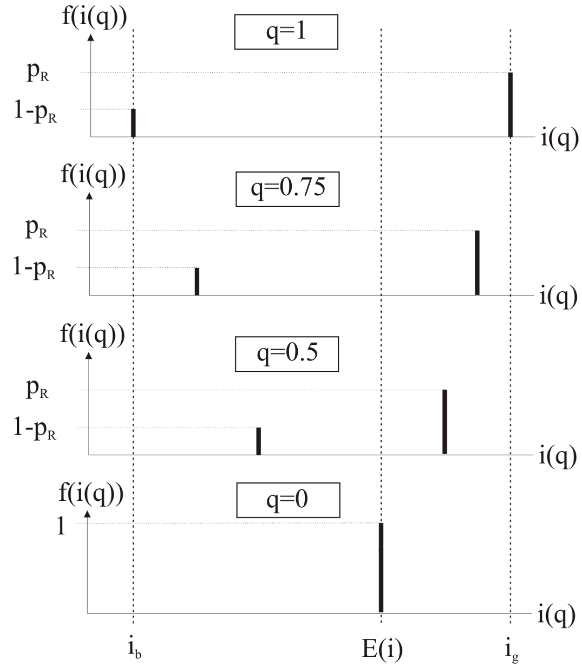


Figure 2.11: Density functions of deposit rates depending on q.

$$E[i(q)] = p_R\left((1-q)E[i] + qi_g\right) + (1-p_R)\left((1-q)E[i] + qi_b\right) = E[i], \tag{2.32}$$

and

$$\begin{aligned} Var[i(q)] &= p_R\left((1-q)E[i] + qi_g - E[i]\right)^2 \\ &+ (1-p_R)\left((1-q)E[i] + qi_b - E[i]\right)^2 \\ &= p_R\left(q(i_g - E[i])\right)^2 + (1-p_R)\left(q(i_b - E[i])\right)^2 = q^2 Var(i), \end{aligned} \tag{2.33}$$

where $Var(i)$ is the variance of the deposit rate if project revenues are perfectly dependent. This is illustrated in Figure 2.11, where we show density functions of the deposit rates, contingent on the realization of q.[71]

Given our binary random variable $\tilde{q}$, we have already shown that an increase in the variance

[71] Note that we used $p_R = \frac{2}{3}$ so that $E[i]$ is closer to i_g than to i_b. The three vertical dashed lines indicate the worst possible bad deposit rate, the expected deposit rate and the best possible good deposit rate, respectively. The actual deposit rates after $\tilde{q}$ has been realized are functions of q and occur where the thick bars are drawn.

at a constant mean is an MPS. Therefore, comparing $i_b(q), i_g(q)$ for $q < 1$ with i_b and i_g (i.e., with $i_b(q)$ and $i_g(q)$ for $q = 1$), the latter is an MPS of the former at any loan rate r. Thus, Rothschild and Stiglitz (1971) applies:

Proposition 2.9 *A higher degree of dependence amongst risky firms as measured by q decreases households' savings at any loan rate r.*

The deposit rates (as functions of r as in Figure 2.3) change as follows. Not all of the properties i) to v) in Subsection 2.2.2 generally apply. The jump of $i_g(q)$ at r_S (property i)) does not need to be upward. For q sufficiently close to 0, it is a downward jump. The bad state deposit rate $i_b(q)$ is monotonically increasing in r in each of the two intervals for $q < 1$ (property ii)). Properties iii), iv) and v) are unaltered. In particular, the expected deposit rate attains its global maximum at r_R and the variance of the deposit rate is monotonically increasing in r. The latter fact can be seen from equation (2.33): $Var(i)$ increases monotonically (property v)) so that $q^2 Var(i)$ does so, too.

The reader can check that the arguments in the proofs of Lemmas 2.1 and 2.2, as well as in Proposition 2.1 and its corollary stay the same. In particular, capital supply and LTU both have their maximum either at r_S or at r_R, irrespective of q. Thus, there are the same six equilibrium cases (cf. Figure 2.6).

The degree of dependence of project revenues might crucially influence equilibrium outcomes in the model. This result is not new, it is only a reformulation of Proposition 2.3. However, the change from independent project revenues to perfectly dependent revenues as in Subsection 2.2.5 is rather extreme.

Proposition 2.10 *There is financial fragility.*[72] *A small change in a parameter can change the type of equilibrium.*

Proof: One such parameter is the degree of revenue dependence as measured by q. Using $p_S = 0.8$, $p_R = 0.2$, $\mathcal{R}_S = 10$, $\mathcal{R}_R = 40$, $N_S = 200$, $N_R = 200$, $C = 2$, $B = 5$, $Y = 2$, $\delta = 0.9$, $H = 1030$, $\theta = 0.40$ and changing q from $q = 0.48$ to $q = 0.49$ decreases capital supply at any given r such that there is a case V equilibrium (with two loan rates) for $q = 0.48$ but a case

[72] See Mankiw (1986), who first characterized financial fragility.

III market-clearing equilibrium for $q = 0.49$.[73]. q.e.d.

2.4.2 Stochastic degree of dependence

Instead of assuming that $\tilde{q}$ has already been realized when households decide about consumption and savings, as we did in the previous subsection, we now assume that $\tilde{q}$ has not been realized yet when they decide. We assume that households know the distribution of $\tilde{q}$. Using the terminology of Knight (1967),[74] there is risk in the degree of dependence, but no uncertainty. In contrast to Subsection 2.4.1, the two state-contingent deposit rates are not known any longer since they depend on the realization of $\tilde{q}$. Therefore, as long as the distribution of $\tilde{q}$ is not degenerate, the deposit rate distribution is not a binary one before $\tilde{q}$ has been realized.

For the consumption-savings decision, we must know the expectation $E[i(\tilde{q})]$ and variance $Var(i(\tilde{q}))$ of i before $\tilde{q}$ has realized.[75] Households take expectations before any random variable has been realized. We get

$$\begin{aligned} E[i(\tilde{q})] &= p_R E\left[(1-\tilde{q})E[i] + \tilde{q}i_g\right] + (1-p_R)E\left[(1-\tilde{q})E[i] + \tilde{q}i_b\right] \\ &= E[i] + E[\tilde{q}]\left(E[i] - E[i]\right) = E[i]. \end{aligned} \tag{2.34}$$

Since $Var(X) = E[X^2] - E[X]^2$, we get the variance by finding the expectation of the squared deposit rate,

[73] Although such a marginal change in a parameter changes the equilibrium case, the allocation is not drastically different. In the given example, the two-price equilibrium has almost no safe firm funded and most of the risky firms get credit at a high loan rate anyway. The difference in the case III equilibrium is that these few safe firms do not get credit and the risky firms have to pay a slightly higher loan rate.

[74] See footnote 25.

[75] This does not mean that we are conducting mean-variance analysis. However, looking at the mean and the variance is sufficient for our purpose as can be seen later on.

$$\begin{aligned}
E[i(\tilde{q})^2] &= p_R E\left[((1-\tilde{q})E[i]+\tilde{q}i_g)^2\right] + (1-p_R)E\left[((1-\tilde{q})E[i]+\tilde{q}i_b)^2\right] \\
&= E[i]^2 + 2E[i]E[\tilde{q}]\left(E[i]-E[i]\right) + E[\tilde{q}^2]Var(i) \\
&= E[i]^2 + E[\tilde{q}^2]Var(i).
\end{aligned} \tag{2.35}$$

As $E[\tilde{q}^2] = Var(\tilde{q}) + E[\tilde{q}]^2$, we get

$$\begin{aligned}
Var[i(\tilde{q})] &= E[i(\tilde{q})^2] - E[i(\tilde{q})]^2 = E[i]^2 + E[\tilde{q}^2]Var(i) - E[i]^2 \\
&= E[\tilde{q}^2]Var(i) = \left(Var(\tilde{q}) + E[\tilde{q}]^2\right)Var(i).
\end{aligned} \tag{2.36}$$

From equations (2.34) and (2.36),

Lemma 2.5 *For all $f(\tilde{q})$ defined on the unit interval, the mean of the distribution of deposit rates is the same and the variance is less than with perfectly dependent project revenues.*

Proof: The first part of the lemma follows immediately from equation (2.34). From equation (2.36), $E[\tilde{q}^2] < 1$ as long as $\tilde{q}$ has support $[0, 1]$, irrespective of the distribution. Therefore, $Var[i(\tilde{q})] < Var(i)$. q.e.d.

Lemma 2.6 *A change in the distribution of $\tilde{q}$ influences savings at a given loan rate r. An increase in either expectation or variance (or both, or any change such that $Var(\tilde{q}) + (E[\tilde{q}])^2$ increases) of the distribution of dependence of project revenues decreases savings, and vice versa.*

Proof: Equation (2.36) implies that the variance of the deposit rate increases due to the changes in the distribution of $\tilde{q}$ indicated in the proposition. Since the mean remains the same, such changes constitute an MPS of the distribution of the deposit rate. Furthermore, since we assumed $\theta < 1$, we know from Rothschild and Stiglitz (1971) that an MPS decreases savings in an expected utility setup with CRRA utility. q.e.d.

We should now go through all the lemmas and propositions in Section 2.2 to check if they still hold. This task is more complex but results in the same consequences for the crucial functions, capital supply and LTU: If $\theta < 1$, savings increase with the well-known discontinuous downward jump, and so does LTU. Again, their maxima occur at the same loan rate, r_S or r_R (just define $z(\tilde{q})$ in an analogous way and apply the same arguments). In the terminology of this subsection, the case of perfect dependence of project revenues can be seen as a special case of stochastic dependence with a degenerate distribution and all probability mass on $q = 1$. It is convenient to resort to the extreme cases of degenerate distributions ($q = 0$ vs. $q = 1$, with probability one each) to see that a change in the distribution of $\tilde{q}$ might crucially influence equilibrium outcomes in our model by changing whether the maximum of capital supply and LTU occurs at r_S or at r_R. Referring back to the transitions in Table 2.1, it is clear that all of these can be caused by a change in the distribution of dependence of project revenues (all other things equal). In particular,

Proposition 2.11 *Given two parameter constellations which differ only in the density $f(\tilde{q})$ of project revenues, it is possible that one leads to an equilibrium with market clearing and the other leads to credit rationing.*

Proof: Consider the two degenerate distributions $\tilde{q} = 0$ and $\tilde{q} = 1$ with certainty (all other parameters equal). The proposition then follows from Proposition 2.3. q.e.d.

Proposition 2.11 also holds for non-degenerate distributions. In fact, all transitions from Subsection 2.2.5 can occur from changes in non-degenerate distributions of the dependence of project revenues. The influence of a change in the distribution on capital supply, LTU and, thus, equilibrium can be illustrated considering the (simple and discrete) distributions in Table 2.2.

The graphs in Figure 2.12 are based on these distributions. The other parameters are: $p_S = 0.6$, $p_R = 0.4$, $\mathcal{R}_S = \frac{40}{3}$, $\mathcal{R}_R = 20$, $N_S = 100$, $N_R = 100$, $C = 2$, $B = 5$, $Y = 2$, $\gamma = 0.8$ and $\delta = 0.9$. On the abscissa appears the loan rate r, and the break-even loan rates are 1.4 and 2.4. The ordinate displays the optimal amount of savings. Going from distribution a)

distribution	q_j	$p(q_j)$	$E[\tilde{q}]$	$Var(\tilde{q})$	$Var(\tilde{q}) + E[\tilde{q}]^2$
a)	0	1	0	0	0
b)	0.5	1	0.5	0	0.25
c)	0.75 0.25	0.5 0.5	0.5	$\frac{1}{16}$	0.3125
d)	0.9 0.5	0.5 0.5	0.7	$\frac{1}{25}$	0.53
e)	1	1	1	0	1

Table 2.2: Various (discrete) distributions of $\tilde{q}$ and some respective characteristics.

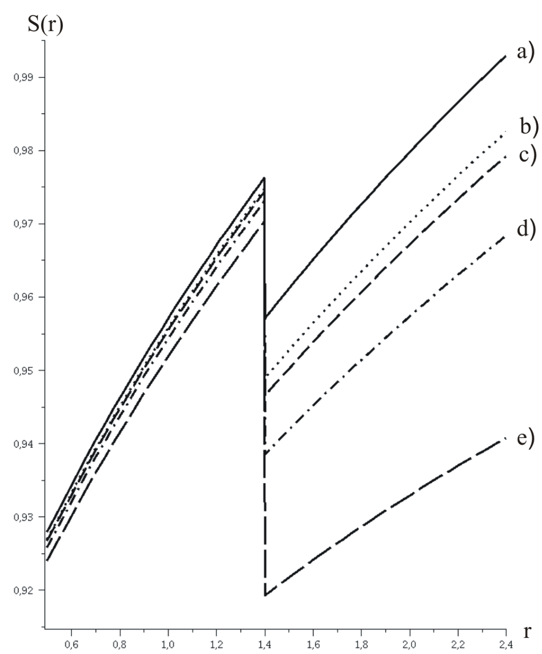


Figure 2.12: Capital supply for distinct distributions of $\tilde{q}$ in one graph.

to e) in Figure 2.12, we can observe what Lemma 2.6 predicts. Any change in expectation and/or variance of the distribution of $\tilde{q}$ that increases the sum of the variance and the squared expectation decreases savings. For distributions a), b) and c), the maximum of capital supply occurs at r_R, whereas distributions d) and e) yield maxima at r_S. LTU also has its maximum at the respective loan rate and the equilibrium outcomes in cases a), b) and c) are different from the equilibrium in cases d) and e).

Having a stochastic $\tilde{q}$ is a useful tool which enables us to go one step further, namely to introduce uncertainty.

2.4.3 Stochastic and uncertain degree of dependence - self-fulfilling expectations

We now assume that households have a common prior about the distribution of the degree of dependence, which might or might not be correct. The correct distribution is unknown, i.e., there is uncertainty (using Knight's terminology, see footnote 25). As in Subsection 2.4.2, households take the consumption-savings decision before $\tilde{q}$ has been realized.

Proposition 2.12 *There can be self-fulfilling expectations: If households expect a high degree of dependence among the risky firms, the equilibrium might be characterized by a high degree of dependence.*

Proof: If the households' prior on $\tilde{q}$ has a very high mean and a very low variance, i.e., households are convinced that there will be a high degree of dependence, savings will be quite low. For a prior with a very low mean and a very low variance, savings could be much higher. As a proof to the proposition, assume the two most extreme cases: a prior of a degenerate distribution of $\tilde{q} = 1$ with probability 1 vs. a prior of $\tilde{q} = 0$ with probability 1. Then, for suitable parameter constellations, transition 8 in Proposition 2.2 can occur,[76] i.e., if households expect a low degree of dependence, the equilibrium could be characterized by market clearing with all projects funded (case VI). By contrast, if they expect a high degree of dependence, the equilibrium can be characterized by market clearing with all risky projects but no safe projects funded (case III).

[76] One such constellation is given in the proof of Proposition 2.2.

We measure the degree of dependence with regard to all firms active in equilibrium, i.e., in relative terms. There is a share $q(1-\beta)$ of firms with (perfectly) dependent revenues in a case VI equilibrium, but a share q of firms with (perfectly) dependent revenues in a case III equilibrium. Since $\beta = \frac{N_S}{N_S+N_R}$, $(1-\beta)q < q$. q.e.d.[77]

2.5 Extension III: intra- and inter-type dependence

Even though we have argued in favor of the assumption that only the risky firms' project revenues are dependent, we show that this is not an indispensable assumption of the model. In this section, we use another sort of dependence. First, we add (perfect) dependence among the safe firms, i.e., in the aggregate, all safe firms will either succeed or fail. Second, we consider inter-type dependence: We assume that risky firms can only succeed if the safe firms succeed.[78] Therefore, the risky firms' success probability is a conditional one. All other assumptions are as in Section 2.2. This section can be considered a robustness test since we show that the main results from Subsection 2.2.5 obtain with these assumptions, too.

The individual success probability of a safe firm does not change, it is still p_S. For the risky firms, we define p'_R as the success probability conditional on the safe firms' success. Let S_j be two Bernoulli random variables which take on the value '1' if all firms of type j succeed, and '0' otherwise ($j \in \{R, S\}$). The four conditional probabilities for the risky firms are: $P(S_R = 1|S_S = 1) = p'_R$, $P(S_R = 1|S_S = 0) = 0$, $P(S_R = 0|S_S = 1) = 1 - p'_R$ and $P(S_R = 0|S_S = 0) = 1$.

[77] Clearly, the degree of dependence as measured by the realization of $\tilde{q}$ is exogenous.

[78] In Arnold, Reeder, and Trepl (2010), this correlation structure is assumed throughout. Major differences to the exhibition in this chapter occur in the definition of a social optimum and in the corresponding welfare analysis.

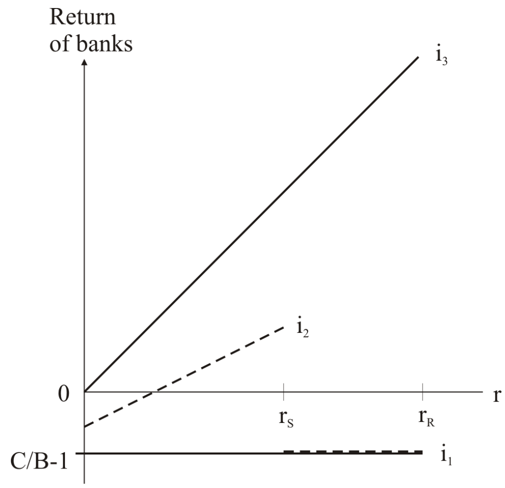


Figure 2.13: State-contingent return function.

The four states of the world[79] occur with the following probabilities:

$$P(S_R = 0 \cap S_S = 0) = P(S_R = 0 | S_S = 0) \cdot P(S_S = 0) = 1 \cdot (1 - p_S) = 1 - p_S,$$

$$P(S_R = 1 \cap S_S = 0) = P(S_R = 1 | S_S = 0) \cdot P(S_S = 0) = 0 \cdot (1 - p_S) = 0,$$

$$P(S_R = 0 \cap S_S = 1) = P(S_R = 0 | S_S = 1) \cdot P(S_S = 1) = (1 - p'_R) \cdot p_S,$$

$$P(S_R = 1 \cap S_S = 1) = P(S_R = 1 | S_S = 1) \cdot P(S_S = 1) = p'_R \cdot p_S.$$

Omitting the zero probability case and defining $p_R \equiv p'_R \cdot p_S$, we have three remaining states.

1. All firms fail. This happens with probability $(1 - p_S)$. The banks' return on lending is $i_1 = \frac{C}{B} - 1$ for all $r \leq r_R$.

2. The safe firms succeed but the risky fail. This happens with probability $(p_S - p_R)$. Returns are $i_2 = \beta r + (1 - \beta)\left(\frac{C}{B} - 1\right)$ for $r \leq r_S$ and $i_2 = \frac{C}{B} - 1$ for $r_S < r \leq r_R$.

3. All firms succeed. This happens with probability p_R. Returns are $i_3 = r$ for all $r \leq r_R$.

Note that p_R gives both the probability that all firms succeed and the unconditional success probability of the risky firms. Figure 2.13 shows a stylized graph of the return function. The worst deposit rate (i_1) is negative and equal to $\frac{C}{B} - 1$, irrespective of the loan rate. The

[79] This time, our use of a 'state of the world' also satisfies the definition of Mas-Colell, Whinston, and Green (1995), cf. footnote 35.

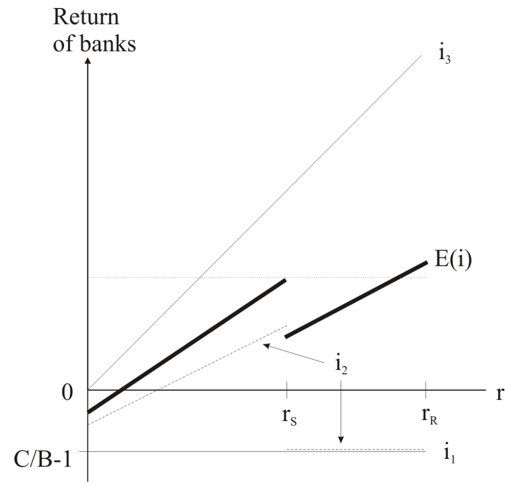


Figure 2.14: Expected return function.

deposit rate in state 2, i_2 is strictly increasing up to r_S, but equal to the worst deposit rate for higher loan rates (since there are no safe firms in the market for loan rates beyond r_S, the rate in the 'safe succeed but risky fail' state equals the rate in the 'all fail' state). The best deposit rate, i_3 is continuous and monotonically increasing in the loan rate with a slope equal to one. The state-contingent returns lead to an expected return function as in Figure 2.14.

In analogy to properties iii) and iv) in Subsection 2.2.2, the expected return function is strictly increasing in each of the two intervals with a discontinuous downward jump at r_S, and the maximum occurs at r_R. A formal proof can be found in Appendix 2.7.4. The intuitive explanation is the same as in Section 2.2, the AR result.

Property v) from Section 2.2 does not hold any more. Even though the variance increases monotonically within each of the intervals, it need not increase at r_S. Thus, it must be asked whether savings still decrease discontinuously at r_S. After all, the lower mean and the lower variance point to opposite directions (for risk-averse households). We argue that savings increase in each of the two intervals for $\theta < 1$, and still decrease discontinuously at r_S.

Let $z^{new} \equiv E[\tilde{R}^{1-\theta}] = (1-p_S)(1+i_1)^{1-\theta} + (p_S - p_R)(1+i_2)^{1-\theta} + p_R(1+i_3)^{1-\theta}$. Then, optimal savings are as in equation (2.8) and LTU is as in equation (2.11) if we replace z by z^{new} in these formulas. Moreover z^{new} increases within each of the intervals if $\theta < 1$, whereas it decreases discontinuously at r_S. This follows from the properties of the state-contingent returns (cf. Figure 2.14) and the definition of z^{new}: Within each interval, the deposit rates weakly increase for all states of the world and raising them to some positive power is a monotonic transformation. At r_S, i_1 does not change, i_3 increases marginally, but i_2 decreases discontinuously, so that z^{new} decreases discontinuously. Since the derivatives of optimal savings and LTU w.r.t. z^{new} are positive for $\theta < 1$, it follows that capital supply and LTU increase within each of the intervals but decrease discontinuously at r_S. In fact, for $r > r_S$, $z = z^{new}$ and, thus, capital supply and LTU are identical to the case where only the risky firms' revenues are dependent. This is because the new assumption implies that there is no change for the risky firms, and because safe firms are not active for $r > r_S$.

Therefore, the proofs regarding shape and maxima of capital supply and LTU are also valid given the dependence structure of this section.

Proposition 2.13 *With intra- and inter-type dependence, all six equilibrium cases from Figure 2.6 are possible. In particular, credit rationing might occur.*

Proof: Plugging the new formula for z^{new} into optimal savings and LTU, the parameter constellations in Table 2.3 in Appendix 2.7.2 again lead to the same six equilibrium cases. q.e.d.

The fact that the equilibrium case is the same under both dependence structures for our chosen parameters does not mean that there is no difference at all. Using $\theta = 0.45$ instead of $\theta = 0.47$ in the first two lines in Table 2.3 in Appendix 2.7.2 leads to different equilibrium cases with the two sorts of dependence. With intra- and inter-type dependence, these are of type II, whereas with only the revenues of the risky firms dependent, the equilibrium cases are of type I.

In general, in a comparison to the case of independent revenues as in SW, both types of dependence lead to lower capital supply and LTU at any given loan rate r (with consequences for the type of equilibrium). For capital supply, this follows from Rothschild and Stiglitz (1971) since the new deposit rate has the same mean but is now risky and, thus, constitutes an MPS of the riskless deposit rate from SW. Comparing our two different dependence structures, we have pointed out that there is no change at all for high loan rates, since only risky firms are active at these rates. For low loan rates, i.e., when both firm types are active, the dependence assumption of this section implies higher aggregate risk. In consequence, deposit rates are more risky such that capital supply at any loan rate r is still lower than with dependence of only the risky firms' revenues.

2.6 Conclusion

Aggregate risk due to the dependence of project revenues has not been analyzed exhaustively in models of the credit market. This is surprising since its consideration can have far-reaching consequences both in a theoretical model and in reality. In a setup similar to the seminal SW model, we have shown that the type of equilibrium can crucially depend on the degree of dependence of project revenues. Since aggregate payoffs become risky, households face capital risk in our setup. Therefore, risk aversion of households becomes a parameter of utmost importance since it influences households' consumption-savings decision. Thus, although the expected deposit rate is maximum at the highest loan rate accepted by borrowers (the AR result), capital supply and LTU do not have to be maximum at that rate. Capital risk deters households from saving so that there might be a credit rationing equilibrium at a lower loan rate. To make this point stronger, we have shown that credit rationing can occur even if a social optimum requires funding of all projects. The definition and analysis of the social optimum significantly differs from a situation with independent project revenues. The new social optimum is characterized both by the number and the type of the projects funded.

We distinguished six different equilibrium cases. We found that dependence of the risky firms' project revenues might reduce the number of safe projects in equilibrium in a socially harmful way. Thus, dependence of project revenues can aggravate adverse selection. Another, non-obvious result in terms of welfare is that the interplay of asymmetric information and dependence of project revenues can lead to an equilibrium with more projects than in the social optimum.

Using a non-expected utility setup to separate aversion against risk per se and aversion against differences in consumption over time, we were able to show that the degree of risk aversion alone is responsible for whether the maximum of capital supply and LTU occur at a lower or higher loan rate. Thus, the equilibrium outcome crucially depends on risk aversion on the one hand and the degree of dependence of project revenues on the other. The latter was made even more clear by showing that leaving aside the restrictive assumption of perfect dependence does not change the conclusion.

As a final robustness check, we showed that our results are not an artifact of our chosen way of modeling dependence of project revenues. In addition to dependence of risky firms' projects, having safe firms' project revenues depending on each other and assuming inter-type dependence at the same time does not change the results.

Our analysis points out that dependence of project revenues is an important factor in the determination of credit market outcomes. We suggest further research on dependence of project revenues in other theoretical models, especially in the area of credit markets. One particular model to analyze is De Meza and Webb (1987), where expected project revenues of safe and risky firms are not the same.

However, aggregate risk has often been neglected or not fully understood in many areas of the theoretical literature in finance. Therefore, we strongly suggest to consider dependencies in areas adjacent to credit markets, too.

In this chapter, we have seen how market frictions can lead to inefficient market outcomes. Asymmetric information can cause credit rationing under certain circumstances. In Section 2.1, we have mentioned some literature which suggests that using collateral to secure loans can mitigate many of these inefficiencies. However, potential borrowers in some economies do not have suitable assets to collateralize loans. Moreover, asymmetric information might not be the most severe market friction. Problems of imperfect enforcement might constitute a much bigger threat to economic efficiency. Given that collateral is scarce, sophisticated contractual arrangements have been proposed as collateral substitutes. Therefore, we consider the performance of such group lending contracts in a market characterized by enforcement problems in the next chapter.

2.7 Appendix

2.7.1 Proof: shape of return function

We successively prove properties i) to v) in Subsection 2.2.2.

i) In the good state, all risky firms and, if they apply for capital in the first place, a proportion p_S of the safe firms succeed. Within both intervals, the composition of the firms with funding is unaffected by a change in r. At r_S, safe firms stop applying for capital so that only risky firms are left in the market in the second interval.

In the first interval, $E[p|g] = \beta p_S + (1-\beta)$ from a bank's point of view. In the second interval, $E[p|g] = 1$. From equation (2.4), an increase in r will increase the expected good state return of a bank by $E[p|g]$ which is strictly positive (and smaller than one) in the first interval and equal to one in the second. Therefore, the good state rate is monotonically increasing within both intervals.

To see that there must be a discontinuous upward jump in the good state rate at r_S, note that, in the good state, banks' returns equal the RHS of equation (2.4) with $E[p|g] = \beta p_S + (1-\beta)(< 1)$ in the first interval. In the second interval, banks' returns are equal to r. The RHS of equation (2.4) is smaller than r for any $E[p|g] < 1$ since $C < (1+r)B$ for any $r > \frac{C}{B} - 1$, and, thus, in particular for $r = r_S$.

ii) In the bad state, none of the risky firms, but a fraction p_S of the safe firms, succeed. In the first interval, $E[p|b]$ equals $\beta p_S(> 0)$, whereas $E[p|b] = 0$ in the second interval. Again, the composition of borrowers does not change within each of the intervals, so that an increase in r will increase expected returns in the bad state by $E[p|b]$. This expression is positive in the first interval, but equal to zero in the second.

To see that the bad state rate is at its global minimum in the second interval, note that banks' returns from equation (2.4) become $\frac{C}{B} - 1$ in the second interval. In the first, at $r = 0$, it must be larger than that since $B > C$, and, thus, any weighted average of C and B (in the numerator of the fraction in equation (2.4)) must be larger than C.

iii) Since the expected rate of return is a probability weighted average of the good and bad state rates, the fact that it is increasing within both intervals follows from i) and ii).

To see that there is a discontinuous downward jump at r_S, consider an infinitesimal increase in r from r_S to $r_S+\epsilon$. All other things equal, this change in r marginally increases the banks' returns. However, since all the safe firms drop out of the market, the average success probability falls discontinuously. While it is equal to $\beta p_S + (1-\beta)p_R$ in the first interval, it equals p_R in the second. Since $p_S > p_R$, the latter expression, p_R, is smaller than the former for ϵ sufficiently small, which proves that expected returns decrease discontinuously at r_S.

iv) The expected rates of return at r_S and r_R are

$$E[i(r_S)] = \frac{(1-\beta)\left(p_R E[\tilde{\mathcal{R}}] + C(p_S - p_R)\right) + \beta E[\tilde{\mathcal{R}}]p_S}{Bp_S} - 1,$$

$$E[i(r_R)] = \frac{E[\tilde{\mathcal{R}}]}{B} - 1.$$

Doing some algebra on these two expressions shows that the former is smaller than the latter iff $E[\tilde{\mathcal{R}}] > C$, which is true by assumption.

v) We prove that the variance increases in each of the two intervals by showing that $(i_g - i_b)$ increases in r. This is sufficient for the proof since

$$\begin{aligned} Var(i) &= p_R(i_g - E[i])^2 + (1-p_R)(i_b - E[i])^2 \\ &= p_R(i_g - (p_R i_g + (1-p_R)i_b))^2 + (1-p_R)(i_b - (p_R i_g + (1-p_R)i_b))^2 \\ &= p_R(1-p_R)(i_g - i_b)^2. \end{aligned} \tag{2.37}$$

Note that these probabilies p_R in the good state and $(1-p_R)$ in the bad state are the same in both intervals. From properties i) and ii), we know that an increase in r increases i_g by $E[p|g] = \beta p_S + (1-\beta)$ and i_b by $E[p|b] = \beta p_S$ in the first interval. Since $\beta p_S + (1-\beta) > \beta p_S$, $i_g - i_b$ increases in the first interval. We further know that an increase in r increases i_g by

$E[p|g] = r$ and i_b by $E[p|b] = 0$ in the second interval wo that $i_g - i_b$ also increases in that interval. Finally, since i_g is increased discontinuously at r_S, whereas i_b is decreased discontinuously, $(i_g - i_b)$ increases in r both within the intervals and at r_S.

2.7.2 Numerical results for social optima

constellation	p_S	$\mathcal{R}_S$	p_R	$\mathcal{R}_R$	N_S	N_R	C	B	Y	H	δ	θ
1	0.8	8	0.2	32	80	220	0.2	6	2	1500	0.9	0.47
2	0.8	8	0.2	32	80	220	0.2	6	2	2600	0.9	0.47
3	0.8	8	0.2	32	80	220	0.2	6	2	2600	0.9	0.40
4	0.8	8	0.2	32	80	220	0.5	6	2	2600	0.9	0.40
5	0.8	10	0.2	40	100	200	2	5	2	1225	0.9	0.30
6	0.8	10	0.2	40	50	250	2	5	2	1600	0.9	0.30
7	0.8	10	0.2	40	200	100	2	5	2	1000	0.9	0.30
8	0.8	10	0.2	40	200	100	2	5	2	1200	0.9	0.30
9	0.8	10	0.2	40	200	100	2	5	2	1000	0.9	0.20
10	0.8	10	0.2	40	120	180	4.5	5	2	1000	0.9	0.25
11	0.8	10	0.2	40	200	100	2	5	2	1500	0.9	0.20
12	0.8	10	0.2	40	200	100	2	5	2	1500	0.9	0.30

Table 2.3: List of parameters.

Table 2.3 shows some parameter specifications which we use to generate Table 2.4. In this latter table, we give the socially optimal levels of investment in both safe and risky firms, as well as the respective equilibrium levels for each of the 12 parameter constellations. We chose parameters in Table 2.3 to get all six equilibrium cases twice: once with the amount of projects funded in equilibrium less than in a social optimum and the other way round, with one exception. By definition of case VI, there cannot be underinvestment since the

case	#all SO	# all E	diff	# safe SO	# safe E	diff	# risky SO	# risky E	diff
I	$127,96$	$122,57$	$-5,38$	$80,00$	$32,69$	$-47,31$	$47,96$	$89,89$	$41,93$
I	$181,74$	$212,46$	$30,72$	$80,00$	$56,65$	$-23,35$	$101,74$	$155,80$	$54,06$
II	$180,22$	$175,29$	$-4,93$	$80,00$	$0,00$	$-80,00$	$100,22$	$175,29$	$75,07$
II	$180,22$	$210,69$	$30,47$	$80,00$	$0,00$	$-80,00$	$100,22$	$210,69$	$110,47$
III	$201,57$	$200,00$	$-1,57$	$100,00$	$0,00$	$-100,00$	$101,57$	$200,00$	$98,43$
III	$217,91$	$250,00$	$32,09$	$50,00$	$0,00$	$-50,00$	$167,91$	$250,00$	$82,09$
IV	$231,16$	$225,20$	$-5,96$	$200,00$	$150,14$	$-49,86$	$31,16$	$75,07$	$43,91$
IV	$254,88$	$270,24$	$15,36$	$200,00$	$180,16$	$-19,84$	$54,88$	$90,08$	$35,20$
V	$264,71$	$250,29$	$-14,42$	$200,00$	$150,29$	$-49,71$	$64,71$	$100,00$	$35,29$
V	$202,27$	$231,89$	$29,62$	$120,00$	$51,89$	$-68,11$	$82,27$	$180,00$	$97,73$
VI	$300,00$	$300,00$	$0,00$	$200,00$	$200,00$	$0,00$	$100,00$	$100,00$	$0,00$
VI	$290,47$	$300,00$	$9,53$	$200,00$	$200,00$	$0,00$	$90,47$	$100,00$	$9,53$

Table 2.4: Inefficiency results.

equilibrium consists of all projects. Thus, there is a zero in line 11, column 4, instead of some negative number.

There are ten columns in Table 2.4. The first one indicates the type of equilibrium (cf. Figure 2.6). The second (fifth, eighth) column gives the total number of (number of safe, number of risky) projects funded in a social optimum, whereas the third (sixth, ninth) column gives the total number (safe, risky) in the respective equilibrium. Column four (seven, ten) is the difference between the former two.

2.7.3 Proof: Lemmas 2.3 and 2.4

The structure of the proofs is identical to the expected utility setup, only using $\hat{R}$ instead of z. Using γ instead of θ in the definition of z, we have $z^{\frac{1}{1-\gamma}} = \hat{R}$.

Proof of Lemma 2.3:

1. If $\alpha < 1$, capital supply increases monotonically in r in each of the intervals $[0, r_S]$ and $(r_S, r_R]$ with a discontinuous downward jump at r_S.
2. Irrespective of α and γ, LTU increases monotonically in r in each of the intervals $[0, r_S]$ and $(r_S, r_R]$ with a discontinuous downward jump at r_S.
3. For $\alpha < 1$, capital supply and LTU have their global maximum at the same loan rate, viz. either r_S or r_R.
4. If s^* is the same at two loan rates, LTU must be the same at these two rates, too.

Proof of 1.

The fact that capital supply increases in each of the two intervals directly follows from Basu and Ghosh (1993, p.124) (Proposition 2). From our equation (2.29), we know that $\frac{ds^*}{d\hat{R}} > 0$ since $\alpha < 1$. The derivative of $\hat{R}$ w.r.t. r is

$$\frac{d\hat{R}}{dr} = \frac{(E[\tilde{R}^{1-\gamma}])^{\frac{\gamma}{1-\gamma}}}{1-\gamma} \left(p_R \frac{1-\gamma}{(1+i_g)^{\gamma}} \frac{di_g}{dr} + (1-p_R) \frac{1-\gamma}{(1+i_b)^{\gamma}} \frac{di_b}{dr} \right), \tag{2.38}$$

which is positive for all $\gamma > 0$. Therefore, for $\alpha < 1$, we have $\frac{ds^*}{d\hat{R}} > 0$ and $\frac{d\hat{R}}{dr} > 0$ so that $\frac{ds^*}{dr} > 0$, too.

To prove that there is a downward jump at r_S, we apply the same argument as in the expected utility case in Lemma 2.1, whose proof was based on Rothschild and Stiglitz (1971). The corresponding argument in the non-EU case was made by Selden (1979): An MPS decreases savings in a non-expected utility setup iff $\alpha < 1$. The fact that there is an MPS at $r > r_S$ together with an increasing capital supply in the second interval completes the proof.

Proof of 2.

Differentiating equation (2.28) w.r.t. $\hat{R}$ and simplifying yields

$$\frac{dLTU}{d\hat{R}} = \frac{Y^{1-\alpha}}{\left(1+\delta^{\frac{1}{\alpha}}\hat{R}^{\frac{1-\alpha}{\alpha}}\right)^{3-\alpha}\alpha\hat{R}}\left[\delta^{\frac{1}{\alpha}}\hat{R}^{\frac{1-\alpha}{\alpha}}\alpha+\delta^{\frac{2}{\alpha}}\hat{R}^{\frac{2-2\alpha}{\alpha}}2\alpha+\delta^{\frac{3}{\alpha}}\hat{R}^{\frac{3-3\alpha}{\alpha}}\alpha\right]. \tag{2.39}$$

Since Y, α, δ and $\hat{R}$ are positive, the derivative itself is positive. Since $\frac{d\hat{R}}{dr} > 0$ in each of the two intervals, $\frac{dLTU}{dr} > 0$ in each of the two intervals. Since $\frac{\Delta\hat{R}}{\Delta r} < 0$ at r_S,[80] $\frac{\Delta LTU}{\Delta r} < 0$ at r_S. Again, the signs do not depend on α or γ. Therefore, LTU increases monotonically in each of the two intervals with a discontinuous downward jump at r_S, irrespective of γ and α.

Proof of 3.

The structure of the proof is as with expected utility: From 1. and 2., and for $\alpha < 1$, the global maximum of LTU and capital supply can only be either at r_S or at r_R. If savings are higher at r_S, $\hat{R}$ must also be higher at r_S since $\frac{ds^*}{d\hat{R}} > 0$ (for $\alpha < 1$), from equation (2.29). Since, from equation (2.39), we also have $\frac{dLTU}{d\hat{R}} > 0$, $LTU(r_S) > LTU(r_R)$. An analogous argument applies if the maximum occurs at r_R.

Proof of 4.

[80] $\hat{R}$ decreases at r_S since capital supply decreases discontinuously at r_S and $\frac{ds^*}{d\hat{R}} > 0$ (for $\alpha < 1$).

The same argument (as for expected utility) applies: If savings are the same at two loan rates, $\hat{R}$ must also be the same. Since LTU in equation (2.28) only depends on s^*, $\hat{R}$ and parameters, LTU must be the same at these two loan rates, too.

Proof of Lemma 2.4:

From equation (2.29), we know that $\frac{ds^*}{d\hat{R}} > 0$ for $\alpha < 1$. Therefore, the maximum of capital supply occurs where $\hat{R}$ is maximum. From the definition of $\hat{R} = (E[\tilde{R}^{1-\gamma}])^{\frac{1}{1-\gamma}}$, we know that the maximum value of $\hat{R}$ (and, thus, whether it occurs at r_S or r_R) does not depend on α. A change in α influences the absolute amount of savings (as can be seen from its appearance in equation (2.26)). However, whether the maximum occurs at r_S or at r_R is independent of α.

2.7.4 Proof: maximum of expected returns for extension III

In the first interval, i.e., for $r \leq r_S$, expected returns are

$$\begin{aligned} E\pi^{bank}(r) &= (1-p_S)\left(\frac{C}{B}-1\right) + (p_S - p_R)\left[\beta r + (1-\beta)\left(\frac{C}{B}-1\right)\right] + p_R r \\ &= (1-x)\left(\frac{C}{B}-1\right) + xr, \end{aligned} \tag{2.40}$$

where the last line uses the definition $x \equiv (1-\beta)p_R + \beta p_S$, $x \in [p_R, p_S]$. In the second interval, i.e., for $r_S < r \leq r_R$, expected returns do not depend on β,

$$E\pi^{bank}(r) = (1-p_R)\left(\frac{C}{B}-1\right) + p_R r. \tag{2.41}$$

Setting $\beta = 0$, we have

$$\begin{aligned} E\pi^{bank}(r_S) &= (1-p_R)\left(\frac{C}{B}-1\right)+p_R r_S \\ &< (1-p_R)\left(\frac{C}{B}-1\right)+p_R r_R = E\pi^{bank}(r_R), \end{aligned} \tag{2.42}$$

since $r_S < r_R$.[81] Setting $\beta = 1$ and using $x = p_S$, we need to plug r_S and r_R from equations (2.2) and (2.3) into equations (2.40) and (2.41), respectively, to see that they are equal.[82] Since expected returns at r_R do not depend on β, the proof is complete if the derivative of expected returns w.r.t. β is positive at r_S. From equation (2.40), we get

$$\frac{dE\pi^{bank}(r)}{d\beta} = -\left(\frac{C}{B}-1\right)\frac{dx}{d\beta} + r\frac{dx}{d\beta} = (p_S - p_R)\left(r+1-\frac{C}{B}\right),$$

for all $r \leq r_S$, which is positive for all positive r since $C < B$.

[81] This makes perfect sense since $\beta = 0$ means that there are only risky firms in the market. Since returns equal project revenue less firm profits, they must be maximum where expected firm profits are minimum, i.e., at r_R.

[82] Which makes sense since $\beta = 1$ means that there are only safe firms in the market such that a loan rate of r_S also extracts all rents from projects.

Chapter 3

Enforcement Problems in Microcredit Markets

This chapter is based on joint work with Susanne Steger and Lutz Arnold. It contains elements of both Arnold, Reeder, and Steger (2009) and Reeder and Steger (2008). The notation is independent of Chapter 2.

3.1 Motivation and the literature

The aim of this chapter is to contribute to the discussion on group lending as opposed to individual lending (IL). Group lending is seen to be one of the main reasons for the success of microfinance, which has in turn been described as one of the most promising attempts to reduce poverty. In a sample of microfinance institutions (MFIs) studied by Cull, Demirgüç-Kunt, and Morduch (2009), 210 out of 315 institutions use some sort of group lending. Giné and Karlan (2009, p.5) distinguish between 'group lending' and 'group liability': "'Group liability' refers to the terms of the actual contract, whereby individuals are both borrowers and simultaneously guarantors of other clients' loans. 'Group lending' merely means there is some group aspect to the process or program, perhaps only logistical, like the sharing of a common meeting time and place to make payments". Our focus is on group lending with joint liability (GL) and if we refer to some other aspect of a group, we explicitly say so.

Cull, Demirgüç-Kunt, and Morduch (2009, p.167) describe microfinance as "a vision of poverty reduction that centers on self-help rather than direct income redistribution". The idea of microfinance is to provide poor households in less developed regions with basic financial services which enable them to start productive activities in order to escape poverty based on their own endeavors.[1] These services include the provision of saving accounts, the possibility to take a loan, and ways to ensure against risks. The literature has labeled these three basic services as 'microsavings', 'microcredit', and 'microinsurance', respectively. The focus of this chapter is on microcredit, i.e., the provision of small-scale loans to poor households.[2] As a starting point into the very young literature on microinsurance, see Morduch (2006) and Chapter 6 in Armendáriz de Aghion and Morduch (2005), where the latter also describes the process of saving in 'household economies' typical of less developed regions.

However, we should note that the conceptual distinction between the three services does

[1] An excellent survey of the circumstances under which the poor make economic choices is given by Banerjee and Duflo (2007). They describe consumption patterns, the way income is generated, and the access to markets and public infrastructure. Before going into the theoretical literature, we recommend the reading of such a survey in order to be able to properly assess the models' assumptions which describe the behavior or environment of the poor. It is mainly the kind of institutions which influence economic choices that are very different from the ones in developed countries. Family, trust, reciprocity and reputation play a much more important role. The literature describes these institutions as 'social capital'. We come back to this point throughout the chapter.

[2] Microfinance is not per se restricted to people in less developed countries (LDCs). Part III in the collection of essays in Carr and Tong (2002) provides three papers on microfinance in the US. Chapter 10 in Yunus (2003) discusses microfinance in the US and in other wealthy countries.

not imply their independence. Among practitioners, it is a widely accepted consensus that the success of microfinance is based on the entirety of financial services. Rhyne (2009) has coined the term ‘inclusive finance’ to stress three facts: First, she notes that the success of microcredit hinges on the interdependence with other financial services. This view is supported by a recent theoretical paper by Ahlin and Jiang (2008), which suggests the complementarity of financial services if the focus is on the long-term success of microcredit. The idea is that the accumulation of wealth depends on the ability to save. The authors find that, although microcredit alone can help break the poverty trap by allowing the poor to harness their productive abilities, microsavings might be crucial to break the ‘mid-income trap’, i.e., to guarantee continuous and lasting income growth. Second, it is not only the poorest of the poor which should be the target of financial service provision. According to Rhyne, there are four billion people who live on less than US$ 3,000 a year (2009, p.3). Third, issues arise not only at the last mile, i.e., between MFIs and borrowers. In the microcredit channel, there are many participants and relationships with conflicts of interest, e.g., MFIs and their loan officers, equity and debt holders of MFIs, public and private investors in MFIs, as well as issues of government involvement. However, it is particularly important to understand the relationship between MFIs and borrowers, on which we focus in this chapter.

The success of microfinance has become known to the wider public when Muhammad Yunus was awarded the Nobel Peace Prize in 2006. More than 30 years ago, Yunus founded the Grameen Bank, one of the world’s largest MFIs, to provide the poor in Bangladesh with capital. Yunus, a professor of economics, questioned conventional wisdom not to lend to ‘the unbankable’. Yunus and his successors proved that collateralizing loans with assets is not a necessary condition for lending. He started lending 27 dollars to 42 people (see Yunus, 2003, p.50). At this stage, Rhyne (2009, p.ix) mentions estimates of the number of active microfinance borrowers “between 60 and 130 million borrowers, depending on who is counting”.[3] We take her latter statement not only as an allusion to special interests in particularly low or high numbers of borrowers, but also to the fact that the microfinance landscape is highly heterogeneous and that there is no canonical definition of an MFI, so that determining the number of (borrowers of) MFIs is subject to some discretion. After

[3]Cull, Demirgüç-Kunt, and Morduch (2009) point to the fact that the number of people concerned is a multiple of this number since most of the borrowers have families.

all, although small scale of a loan is frequently taken to indicate a microcredit, there is no clear-cut threshold in terms of loan size. The average loan balance of Badan Kredit Desa of Indonesia is \$38 (see Morduch, 2000, p.618), whereas others, also labeled MFIs, give average loans of several thousand dollars (Bolivia's BancoSol, for instance).

Apart from differences in the size of loans given, there are large variations in political, geographical, economic and social settings. First, MFIs differ in terms of their legal status and profit status. Cull, Demirgüç-Kunt, and Morduch (2009, p.174) use five different categories: banks, NGOs, nonbank financial institutions (NBFIs), credit unions, and rural banks (75 % of the institutions in their sample are either NGOs or NBFIs). Second, MFIs operate in very different geographic areas all over the world. Clearly, conditions in Bangladesh differ from the ones in sub-Saharan Africa, Eastern Europe, South America or Asia. On a country level, some are frequently hit by natural disasters, whereas others are located in rather stable climatic regions. Moreover, cultural habits, religion, and values and virtues strongly differ between regions. Third, intentions and use of microloans can differ considerably. Ahlin and Jiang (2008, p.1) stress the investment character of microloans, defining them as "small amounts of capital (...) to facilitate income-generating self-employment activities". By contrast, some MFIs also give loans for urgent consumption needs. A directly connected feature is that some MFIs commit the disbursement of loans to particular purposes. This points to, fourth, differences in the lender-borrower relationship. Some MFIs restrict to operative banking functions like procurement of funds, disbursement of loans, management of staff and collection of interest payments, and appropriate enforcement techniques. Others accompany borrowers throughout the process of investing and even provide them with basic education and advice.

This heterogeneity gives a 'raison d'être' for the immense amount of country studies sometimes classified as being part of the rather descriptive 'development practitioner's literature'. It also questions the 'best practice' approach frequently pursued by international donor institutions.

From basic economic theory, one can question why capital does not 'naturally' flow to the poor, as Armendáriz de Aghion and Morduch (2005) do in their Section 1.2. The law of diminishing marginal returns to capital would predict that the least endowed individuals are the most productive so that investors should be expected to compete for the privilege to serve the poorest of the poor. Of course, this comparison is flawed since the law has been formulated on ceteris paribus grounds, i.e., holding everything else equal. In fact, infrastructure and human capital (and many other factors that influence productivity) are highly unequal between and also within countries. Not only do little infrastructure and low levels of education influence marginal productivity, they also increase the cost of lending.

This cost can be divided into operating costs, capital costs and loan loss provisions. According to Cull, Demirgüç-Kunt, and Morduch (2009, p.183), capital costs and loan loss provisions are rather constant among the five categories of MFIs mentioned above, but operating costs differ considerably (cf. their Figure 2). This is not a big surprise given the above mentioned heterogeneity of the microfinance landscape. Operating costs arise from administration as well as from risk appraisal, where the latter is directly connected to the information asymmetries lenders face, and the difficulties to deal with them. However, the most important cost factor is the lack of scale economies, which might arise from two different sources. On the one hand, the cost of lending tends to decrease as the MFI grows larger. On the other hand, the size of loans reduces the relative cost per loan. A main result of Gonzalez (2007, p.39) is that the cost reduction resulting from an MFI's scale disappears from 2,000 borrowers on. Thus, it is the small size of loans which crucially causes operating cost and, thus, total transaction cost to be much higher than in traditional economies.

High transaction costs have frequently been used to justify subsidies for MFIs.[4] Morduch (2000) explains the 'subsidy trap'[5] and discusses some cases of conventional wisdom with respect to subsidies. For instance, it is appealing to think that subsidies reduce efficiency and, thus, profitability. However, Morduch points out that aiming at profits is a sufficient but not necessary condition for efficiency. It is more important to have hard budget constraints.[6] For the purpose of our analysis later in this chapter, the most important claim Morduch invalidates is that financial self-sufficiency, which only few MFIs exhibit, is necessary to attract commercial funds. He concludes that the "chief constraint is not subsidization *per se*, but

[4]Subsidies come from various sources and at different conditions, ranging from NGO donations to government loans at preferential loan rates.

[5]He describes the problem as a vicious circle. Poorly managed subsidized credit programs imply exorbitant default rates. This releases borrowers from the shame of defaulting since everybody else is doing so. By receiving ample fresh capital from governments to cover losses, incentives to mobilize savings are low, too, which weakens the process of development further since it is the entirety of financial services that fosters steady income growth. Clearly, incentives to improve on efficiency are also low in these programs.

[6]Other commonly held claims Morduch puts into perspective include the fear that subsidized credit ends up in the hands of the non-poor, that government involvement is detrimental, and that subsidies limit savings mobilization.

the ability to limit perceived riskiness" (p.623). Thus, it seems reasonable to assume that the majority of MFIs turns to global capital markets for funding, even though very few are financially sustainable. Another justification for subsidies is based on outreach and impact of the supported MFIs. The idea is that some of the poorest borrowers cannot be served in a financially sustainable manner so that subsidies are necessary to provide these people with access to financial services. In Chapter 4, we will have much more to say about the trade-off between financial returns and outreach.

Before analyzing how MFIs channel funds to their borrowers, in particular by determining which of the two lending types (IL vs. GL) occurs in equilibrium, we want to briefly review other factors which have been assigned crucial roles for the success of microfinance, and whose relative importance is subject to significant debate. They all have an effect on the cost of lending, either by reducing operating costs or by reducing loan default through increased repayment rates.

Operating costs are reduced by the introduction of new management techniques, as for instance the requirement that loan disbursement and repayment take place at some central venue. Borrowers are expected to assemble there, so that a loan officer saves traveling cost and time.[7] Another frequently mentioned fact, especially for the high repayment rates, is that most MFIs focus on female customers, which are considered more reliable and diligent borrowers.

Other features provide MFIs with information otherwise unavailable. For instance, MFIs rely on tough repayment schedules, some requiring borrowers to make small repayments every week. Thus, loan officers can quickly build credit histories and learn about potential risks. These risks are further reduced by 'progressive lending', i.e., by providing borrowers with small loans initially and only handing over larger sums in case borrowers prove their reliability. Moreover, weekly meetings allow loan officers to establish personal relationships to borrowers. All these features can significantly increase repayment rates.

A central issue for lenders in developing countries is the problem of collateral, which is related to the problems of enforcement. Limited wealth is certainly the main obstacle for collateralization, sometimes due to limited property rights. However, even if there were

[7]However, readers familiar with the traveling salesman problem would agree that this shift of effort from loan officers to borrowers will have a negative effect in overall distance, although it certainly decreases the bank's cost.

suitable assets to use, poorly functioning legal systems make enforcement one of the main problems. Bond and Rai (2002) discuss two approaches to mitigate the problem: dynamic incentives and social sanctions. The former refers to the threat of future credit denial in case of default, the latter to the use of 'peer sanctioning' based on the existence of 'social capital'. For instance, repayment is generally done in public which gives borrowers incentives to work hard in order to avoid social stigmatization. Putnam (1993) defines social capital as "features of social organization, such as networks, norms, and trust, that facilitate coordination and cooperation for mutual benefit". In the realm of microcredit, social capital mainly refers to three facts within close-knit communities. In comparison to industrialized societies, people living in such communities are assumed to know more about each other, to be in a better position to observe actions of their peers, and to have more powerful, informal sanctioning mechanisms, i.e., the cost of peer screening, peer monitoring, and peer sanctioning is low in developing countries. In our formal model, we start from a situation without social sanctions and then see their crucial influence on equilibrium.

Before that, we review some of the literature more closely related to the effects of joint liability in groups, which has mainly taken place in the strongly interconnected subfields of information economics, contract theory, and mechanism design.[8] When the GL contracts of Grameen Bank and other MFIs became widely known to the academic community, the literature on contract theory had extended principal-agent models to consider principals facing several agents. These multi-agent environments consider a wide range of contracts and ask which one is optimal. Thus, the contracting problem can be divided in two steps: the description of the set of available contracts and the choice of the optimal contract. Typically, endogenous parameters in the choice of an optimal contract include the loan rate, the amount of collateral (if there is some), the penalty in case of default, and the degree of joint liability. When a principal faces several agents, possible interactions between agents are of utmost importance to design incentive compatible contracts. In the terminology of contract theory, the degree of possible side trades (or side contracts) matters a lot. Holmström and Milgrom (1990, p.335) define side trades as "implicit or explicit exchanges between the agents which the principal cannot control directly because he cannot observe them". They further specify

[8]Most of the current literature on contracts and mechanism design is based on the assumption of asymmetric information.

two key sorts of side trades: collusion and cooperation. Whereas collusion are trades that harm the organization (which the principal and agents are assumed to belong to), cooperation are trades that help the organization. However, they also admit that "very similar trades can be labeled as cooperation or collusion depending on the context and the interest of the trade" (p.335). Most of the more recent papers we present below find optimal contracts given one or another form of side contracting.

However, let us start with some results based on models which try to explain *how* GL can be beneficial at all, as opposed to the mechanism design literature which determines how contracts can be *most* beneficial. In terms of mitigating problems arising from asymmetric information, GL can help to reduce adverse selection and moral hazard.

In the realm of adverse selection models, Armendáriz de Aghion and Gollier (2000) consider two channels through which GL can increase efficiency, a 'collateral effect' and a 'self-selection effect'. In their model, inefficiencies arise in the case of IL contracts if cross-subsidization from safe to risky borrowers discourages safe borrowers with socially desirable projects from applying for credit. GL can restore efficiency by lowering equilibrium loan rates. In both cases, there is asymmetric information in that the bank cannot observe the types of borrowers it faces. Borrowers are either safe or risky and are protected by limited liability. The collateral effect can work even if borrowers do not know each other's type, which implies the existence of groups with borrowers of the same type, as well as with different types. Risky borrowers have higher payoffs if they succeed so that they can always shoulder a partner's default if they succeed themselves. By contrast, safe borrowers are assumed to have payoffs insufficient to repay the whole group loan. Thus, by limited liability, the externality from a risky partner's default is not fully borne by a safe borrower. GL reduces the probability of default and can then lead to lower equilibrium loan rates such that both types of borrowers apply for loans (see the presentation in Armendáriz de Aghion and Morduch, 2005, pp.94-96). Efficiency can be restored without relying to the informational advantage poor borrowers are usually assumed to have over lenders. By assuming that borrowers know each other's type, Armendáriz de Aghion and Gollier (2000) show the self-selection effect (also known as 'assortative matching') of GL. Safe borrowers are shown to match with safe borrowers so that the risky will have to form groups among each other. In that case, even though banks cannot observe the types in a group, GL might restore efficiency by the equilibrium loan rate channel again: Whereas cross-subsidization deters safe borrowers from lending with IL contracts, joint liability leads to lower equilibrium loan rates for both types. Note that, even though both types of groups pay the same nominal loan rate, there is no cross-subsidization since safe borrowers pay lower effective loan rates - they have to repay for their partners less often.

Ghatak (1999) also derives the self-selection effect and proves that the assortative matching result survives if side payments are allowed. In Ghatak (2000), he shows an interesting consequence of assortative matching, namely that banks can use GL as a screening device. The separating equilibrium entails two kinds of contracts,[9] one with a low loan rate but a high degree of joint liability, and the other with high loan rates but a low degree of joint liability. Risky groups would choose the latter, whereas safe borrowers would opt for the former since the probability that the joint liability clause applies is low.[10] Thus, a first-best allocation can be achieved, charging risk-adjusted rates to both types. Van Tassel (1999) has shown another case of screening by joint liability. He derives a slightly different result: If borrowers can choose between IL and GL contracts, safe borrowers match and choose group contracts. Risky borrowers could then group with each other and choose group contracts, but opt for IL contracts instead.

In another adverse selection model, Laffont and N'Guessan (2000) point out the importance of social ties. If borrowers do not know each other, the collateral effect of joint liability does not work.[11] If they do know each other, group lending contracts are shown to be efficient. Varying the possible extent of side trades, simple GL contracts are not collusion-proof. The authors also provide a description of contracts which are robust to collusion.

One of the early papers on the effects of GL on moral hazard is Stiglitz (1990). It is one of the four models which Ahlin and Townsend (2007) in terms of their predictive power for repayment rates, the other models being Banerjee, Besley, and Guinnane (1994), Besley and Coate (1995, BC) and the above mentioned Ghatak (1999). Ahlin and Townsend choose these four models since they are among the most widely cited. Stiglitz (1990) describes 'peer monitoring' among borrowers in a group in a situation with problems of ex ante moral hazard

[9] Cf. the concept of separating equilibria in Bester (1985) and Stiglitz (1977).

[10] Safe borrowers having selected safe peers need to stand in for their group members only if their partner defaults and they are successful themselves.

[11] This contradicts Armendáriz de Aghion and Gollier (2000). However, comparing both models, the assumptions necessary to derive the collateral effect in Armendáriz de Aghion and Gollier (2000) seem rather artificial.

with side-contracting. By joint liability, borrowers suffer from default of a group member so that they have the proper incentives to monitor each other's efforts. Via higher repayment rates, equilibrium loan rates can be reduced. In Stiglitz' model, GL also has a negative effect. It transfers risk from the bank to the cosigner. However, in the type of equilibrium considered, Stiglitz shows that the overall effect is positive.

Another ex ante moral hazard problem is analyzed by Varian (1990). He considers a principal facing several agents in a setup where information is costly. The crucial assumption is that agents can obtain information about (the actions of) their peers at a lower cost than the principal, as is common for village economies targeted by MFIs. Varian does not consider joint liability GL, but focuses on the nature of supervision delegated to agents, as well as on the desirability of side trades from the principal's point of view. The side trades considered include optimal supervision and mutual insurance.[12]

Banerjee, Besley, and Guinnane (1994) set up an optimal contract framework to consider IL contracts with a cosigner. Similar as in Stiglitz (1990), there is an ex ante moral hazard problem. Borrowers are protected by limited liability, so that they are tempted to undertake riskier projects. In contrast to Stiglitz, monitoring of the borrower is costly to the cosigner and a prerequisite for punishment. The endogenous parameters are the degree of internal project funding, the degree of the cosigner's liability and the loan rate. They find that the borrower's incentives to take risks can be reduced since the cosigner's liability makes him threaten the borrower with punishment.

Other models have focused on enforcement problems, which are particularly present in developing countries and constitute a major obstacle for welfare-enhancing trade. The standard reference is the model of BC.[13] It is one of the few models which do not use contract theory. Incomplete enforcement is expressed by the lender's inability to enforce contractual claims. Thus, if a bank anticipates that it cannot enforce repayment from borrowers, even if the latters' projects are highly profitable, it might not consider lending funds. BC investigate the impact of joint liability in borrower groups on loan repayment rates.[14] For that purpose, they set up the 'repayment game', in which borrowers non-cooperatively decide about whether

[12] He also points to group formation issues as treated in more detail by Armendáriz de Aghion and Gollier (2000), and Ghatak (1999, 2000), see above.

[13] As of 05.12.2009, Google Scholar gives 554 citations.

[14] See also Ghatak and Guinnane (1999, p.209) and Armendáriz de Aghion and Morduch (2005, pp.297-298).

to repay or not. Returns are sufficient to repay with certainty in their model, but due to enforcement problems, borrowers do not repay unless the penalty for default weighs heavier than the burden of repayment. GL, as opposed to IL, has two effects on repayment rates then. For one thing, it enhances repayment when one borrower is able and willing to stand in for a member of a group who does not repay. For another, liability for the repayment of the other group member potentially discourages a borrower from repaying at all when he would have repaid an individual loan. BC's first key result is that, when payoffs are independently and uniformly distributed and penalties for default are proportional to payoffs, GL leads to a higher repayment rate when the loan rate is sufficiently low, while the repayment rate is higher with IL when the loan rate is high. Their second main result is that social sanctions imposed on group members who refrain from making their contribution to the contractual repayment reduce the drawback of GL. If the sanctions are severe enough, GL always (i.e., for all loan rates) yields the higher repayment rate.

These are important results which shed much light on the important issue of repayment rates given non-cooperative behavior of borrowers. However, as called for by Townsend (2003), the recent literature tends to analyze microcredit in mechanism design frameworks.

Laffont (2003) analyzes the effects of correlated project returns in an adverse selection model with limited liability. In particular, he analyzes the performance of GL contracts in different classes of contracts and finds positive effects when returns are correlated but points out that GL contracts are not necessarily optimal, depending on the degree of possible side trades and the class of mechanisms considered.

Laffont and Rey (2003) challenge the claim that banks benefit from collusion between borrowers when the latter observe each other's effort. The authors separate information sharing from collusion. They point out that information sharing is good, even if it entails collusion, but that collusion per se is bad.

Rai and Sjöström (2004) analyze a model with ex post moral hazard, i.e., with asymmetric information regarding project revenues. Since banks cannot observe borrowers' project revenues, borrowers have incentives for strategic default. Inefficiencies arise from (non-pecuniary) penalties banks can impose. Thus, the optimal contract minimizes expected punishments. From the assumptions of the model, it follows quite trivially that complete insurance (side) contracts are optimal when ex ante side contracting is possible. In such a situation, IL contracts are efficient so that there is no room for improvement using group lending contracts. However, assuming that only interim side payments are possible, they exclude state-contingent side contracts and, thus, efficiency of individual loans. In their analysis of group lending, the authors distinguish between two group contracts: simple joint liability (i.e., GL) contracts and cross-reporting mechanisms. While GL contracts induce mutual insurance, they are not efficient since the deadweight loss of a high penalty applies if the projects of both borrowers fail. The authors are able to show how cross-reporting[15] achieves a reduction of punishment if both borrowers' projects go awry without influencing repayment incentives in the other cases.

In an interesting contribution, Conning (2005) assumes a monitoring technology which both peers and delegates can use. In a framework with moral hazard and limited liability, he compares three lending mechanisms. Lending via (collateralized) IL contracts, GL, and IL accompanied by delegated monitoring which is subject to moral hazard, too. He finds an advantage of GL which rests on an 'incentive diversification effect' when compared to a delegated monitoring agreement. With the latter lending type, a lender has to cede two rents to ensure proper incentives for both agents, whereas the former allows to economize on the cost of aligning incentives.

The starting point for Bhole and Ogden (2009) are several papers finding a negative influence of GL under certain circumstances. We discuss some arguments against GL before presenting their results. We have already mentioned Stiglitz' assertion that joint liability transfers risk from banks to borrowers. Furthermore, Giné and Karlan (2009, p.5) describe how joint liability can cause different kinds of tensions between borrowers in a group. Social sanctions and peer monitoring, the panaceas in terms of economic theory, can have serious negative consequences for interpersonal institutions like trust or reciprocity. The same is true for free-riding, which occurs when a successful borrower does not contribute to the group repayment expecting other members to shoulder his debt. The effect is explained in detail in our formal model below. Over time, GL might face further difficulties. If new members with looser social ties enter (cf. the result of Laffont and N'Guessan, 2000), GL might perform worse. Moreover, changes in individual credit demand over time can lead to diversion of credit needs. Borrowers with smaller loans dislike liability for peers with larger loans. In terms of repayment rates, we have already explained the result of BC, who find that joint liability can

[15] A group member sends a report specifying his partner's repayment ability to the bank.

discourage a borrower from repaying at all when he would have repaid an individual loan.

Another motivation for the Bhole and Ogden (2009) paper is dissent with the cross-reporting mechanisms in Rai and Sjöström (2004), in particular since they are not widely observed in reality. In an attempt to explain why such mechanisms are not observed, we concur with Bhole and Ogden in that these messages could create tensions between borrowers. As a consequence, borrowers might want to keep as much of their private information, thus inhibiting knowledge spillovers. Far worse, such systems which implicitly suggest that borrowers shirk and cheat might destroy one of the most important resource of the poor, social capital. In order to get incentives right by creating sophisticated extrinsic reward and punishment systems, intrinsic motivation of borrowers might be destroyed. We further comment on this problem in the conclusion of this chapter.

Bhole and Ogden consider three crucial endogenous variables of group lending contracts: the degree of joint liability, the possibility to renegotiate on repayments and a non-pecuniary penalty. They assume asymmetric information and enforcement problems, i.e., banks do not observe project revenues and cannot extract payments from borrowers if borrowers are unwilling to repay.[16] The main contribution of Bhole and Ogden (2009) is to show how contractual deficiencies of simple GL contracts as in BC and other papers can make GL worse than IL.

The results of the mechanism design literature are important in that they describe how contracts should be designed in order to maximize borrower welfare. In our equilibrium analysis, we also use borrower welfare as the crucial determinant of the equilibrium contract. Since we assume perfect competition, our analysis is similar to an optimal contract approach with two endogenous variables, lending type and loan rate. Before we explain in detail what we do and find, we give a motivation for our approach, in particular in light of the results of the mechanism design literature.

Although the literature is well aware of the missing equilibrium analysis of the BC model,[17] our analysis as it is (i.e., with an exogenous, proportional penalty function) shows that repayment rates are not a good indicator to predict equilibrium outcomes. We also apply an

[16]It is not unambiguously clear if there are enforcement problems in their model. On the one hand, banks cannot 'directly' extract payments from borrowers. On the other hand, the non-pecuniary penalty a bank can impose is so high that a borrower prefers to pay back all of his project revenue in order to avoid the penalty. We further comment on this fact below.

[17]See footnote 2 in Ahlin and Townsend (2002, p.6).

analysis of the allocation problems from the imperfect information literature to microcredit markets with enforcement problems.

We acknowledge the power of mechanism design frameworks compared to an equilibrium analysis. In a market model with perfect competition, there is not much need to set up an equilibrium model. In fact, as Hart and Holmström (1987, p.74) argue, considering a problem of optimal contracts has advantages over an equilibrium analysis since "[m]ethods for solving optimization exercises are substantially more advanced than methods for solving equilibrium models". In particular, "the fact that market forces reduce to simple constraints on expected utilities greatly facilitates equilibrium analysis".

Conning (2005) considers a large variety of circumstances in his mechanism design approach and explains the merits of his model in terms of an equilibrium analysis as follows: "Putting all of the variants of the model together ends up providing a rich set of predictions regarding the shape of market structures that may emerge on a competitive loan market with heterogeneous borrowers who differ in terms of initial collateral asset holdings and project characteristics and how financial intermediaries expand and transform the set of trades that can take place. Some borrowers are offered, and will choose, joint liability loans, other may prefer individual liability contracts with or without delegated monitors. Yet others will remain excluded from the loan market" (p.5).

In our formal model in this chapter, we analyze heterogeneity regarding the degree of the enforcement problem in Subsection 3.4.5. Our result on redlining confirms his statement.

As powerful as mechanism design is, Hart and Holmström (1987) also recognize the limits of the theory of optimal contracts in case these contracts become "monstrous state-contingent prescriptions" (p.74). They also note that "substituting an optimization analysis for an equilibrium analysis is not always economically meaningful". We argue that our equilibrium analysis of the BC model does make sense even though the contracts involved (IL, simple GL, and GL with social sanctions) are rather restrictive, in particular due to an exogenous penalty function and the requirement that repayment is an all-or-nothing decision.

Both Bhole and Ogden (2009) and Rai and Sjöström (2004) do not have to put up with criticism that inefficiencies arise from non-optimal contracts. However, any optimal contract framework must make assumptions about the specific form of these penalties. In a way, endogenizing the penalty function thus shifts optimality to the discretion of the economist who sets up the model. For instance, Bhole and Ogden (2009) assume binary penalties which come about randomly, i.e., the penalty occurs with a certain probability and no punishment occurs otherwise. This might not be feasible in reality. Borrowers would probably doubt that it was more than chance which freed their neighbor from penalty but imposed a major penalty on themselves. Thus, it is not unambiguously clear that an endogenous penalty function is the superior assumption. Moreover, penalties in both of the aforementioned papers can be very high. In fact, both assume that penalties can be higher than the best possible project payoff, which is at odds with our understanding of an enforcement problem. Furthermore, both assume that penalties are completely non-pecuniary. In our analysis, we also look at the effects of a penalty which is in part pecuniary.

Another argument in favor of an exogenous penalty function is based on Conning (2005, p.2) who finds that "the group methodology is often rigid and poorly adapted to borrowers' changing needs". This supports considering group lending in its purest form and to avoid complex contractual structures as in Rai and Sjöström (2004). Mechanism design results are important to show how contracts should be shaped, but as long as contracts are not what they should be, an equilibrium analysis given some actually used contracts might make much sense.

As another advantage of our model, and in contrast to Bhole and Ogden (2009) and Rai and Sjöström (2004) who assume that projects either succeed or fail, we assume that project revenues are distributed on some interval. Thus, we can derive a more realistic return function to describe the kind of market failures known from the asymmetric information literature.

The focus on enforcement problems in our model is supported by Giné and Karlan (2009), who find "that peer monitoring or peer pressure are unimportant [for repayment]". Moreover, there is some empirical evidence that is supportive of the BC model and its basic mechanisms.

Ahlin and Townsend (2007, p.F42) study the Thai BAAC and find that "the Besley and Coate model of social sanctions that prevent strategic default performs remarkably well, especially in the low-infrastructure northeast region". Their more general finding that no single model is able to capture the observed diversity of different microcredit programs, may also help explain why the BC model fares worse in other studies.[18]

BC themselves recognize that their results "should not be taken as implying that group

[18] For instance, see Cassar, Crowley, and Wydick (2007, p.F101), and Giné and Karlan (2009, p.13).

lending is better or worse than individual lending in any broader sense than repayment rates" (p. 16), so "a more comprehensive analysis of the differences between the two lending schemes is an interesting subject for further research" (p. 16). This "requires a richer framework than provided by the model of this paper" (p. 15). The aim of the present paper is to supplement the BC model with a minimum set of additional assumptions that enables us to study the nature of equilibrium and the welfare effects of the different lending schemes.

The analysis is further motivated by the recent huge inflow of private investors and the surge in the use of market instruments in the market for funding microfinance institutions (MFIs). According to Reille and Forster (2008, p.1), "[t]he entry of private investors is the most notable change in the microfinance investment marketplace. (...) Individuals and institutional investors – including international retail banks, investment banks, pension funds, and private equity funds – are all looking for ways to channel capital into microfinance, and investment banking techniques are being introduced to create investment vehicle alternatives that appeal to an increasingly broad range of investors." The total volume of the microfinance market is estimated at US$ 25 billion in 2006. According to Dieckmann (2007, pp.6-7), the top 150 MFIs are by now mature, mostly regulated, and profitable institutions, and a further 800 are set to follow. These MFIs increasingly attract funds from individual and institutional investors, other sources of funds being development finance institutions (DFIs).[19] In particular, private investors make up for more than 50 percent of the US$ 4 billion foreign investment in MFIs. A similar percentage of the cross-border investments is made not directly but via specialized microfinance investment vehicles (MIVs). Except blended-value funds (BVFs), Reille and Forster (2008, Figure 1, p.2, and Table 1, p.7) find that most MIVs meet return expectations of about 5 percent. Investment banks have started to securitize MFI claims in the form of CDOs. Dieckmann (2007, p.10) suggests a back-of-the-envelope calculation that highlights the vast growth potential of the microcredit market: "While MFIs currently serve an estimated 100 million micro-borrowers, the total potential demand is roughly estimated at 1 bn" (given low penetration rates of below 3 percent in large markets such as India and Brazil). So this US$ 25 billion market may grow ten-fold if it attracts the required funds. Exhausting this growth potential necessitates a continuation of private capital inflows attracted by decent returns. So there is little doubt that the recent trend toward private investments

[19] Our definitions of MFIs and DFIs follow Reille and Forster (2008).

and market instruments will continue over the foreseeable future.

The equilibrium analysis of the BC model brings forth several interesting results. To begin with, we consider the model without social sanctions. The equilibrium loan contract maximizes expected borrower utility subject to the constraint that the MFIs break even. We demonstrate that an equilibrium exists. In equilibrium, the borrowers get the finance needed for their projects or not, depending on the level of expected returns and the nature and severity of the penalties. Our first main result is that in an equilibrium with IL, it may be possible for MFIs to offer a GL contract that has a lower loan rate, increases the repayment rate, and breaks even – but no borrower accepts this low-interest offer. The reason is that GL with non-cooperative behavior in the repayment game of BC entails potentially large penalties, which are not only threatened, but also exerted. Theoretically, this shows that repayment rates are an imperfect indicator for the viability of lending types in equilibrium, systematically biased toward GL.

Furthermore, we find that the return function (that relates the return on lending to the loan rate) is a hump-shaped function over the relevant range of loan rates, so that there are different kinds of market failure. The equilibrium may be characterized by financial fragility, in that a small increase in the rate of return required by the investors (viz., a rise beyond the maximum of the return function) leads to a complete breakdown of the market (cf. Mankiw, 1986). From a cross-sectional perspective, when there are several microcredit markets of the BC type, redlining as in Riley (1987) may occur: All borrowers get loans in some markets (those where the maximum of the return function is no less than the rate of return required by the investors), while no-one gets credit in other markets. This result may be helpful in understanding why microfinance works well in some places but not in others. It implies that in order to maximize the total volume of credit given, DFIs should target the least profitable markets consistent with their return expectations, leaving the more profitable segments to private investors. Credit rationing may also arise: In a given market, some borrowers get funds, while other, indistinguishable borrowers do not. In sum, our second main result is that, irrespective of whether GL or individual lending arises as the equilibrium mode of finance, microcredit markets with problems of enforcing repayments are likely to be characterized by the usual types of allocation problems encountered in loan markets with asymmetric information.

We also consider the case of cooperative behavior of borrowers. As expected, the counter-intuitive result that IL might be the equilibrium lending type even though GL has a higher repayment rate and a lower equilibrium loan rate disappears. We are able to show that repayment rates and expected repayments are higher with GL than with IL when borrowers in groups cooperate and the banks' penalties are non-pecuniary. We show that GL becomes the unique equilibrium mode of finance.

We also analyze the model with social sanctions and are able to show that GL is the unique equilibrium mode of finance when borrowers do not cooperate in the repayment game but apply (sufficiently large) appropriate sanctions to each other. Interestingly, since we assume that borrowers are always able to repay loans by selling their belongings, GL with social sanctions can have negative effects. It might force borrowers to sell their belongings and repay in order to avoid social sanctions. In terms of the market equilibrium, borrowers do not demand loans at any loan rate since their utility from doing so can be negative. However, when the banks' penalties are non-pecuniary, all loan rates that banks subject to competition could offer attract borrowers if borrowers are risk-neutral.

The allocation problems identified in the model without sanctions persist. That is, social sanctions ameliorate, but do not eliminate, the negative impact of enforcement problems on equilibria in microcredit markets. In terms of the power of joint liability, we conclude that there need to be other factors to guarantee the success of group lending.

The remainder of the chapter is organized as follows. Section 3.2 describes the model without social sanctions. Section 3.3 summarizes BC's results on repayment rates. In Section 3.4, we go on to characterize the model equilibrium (details of the derivations are delegated to the appendix of this chapter, 3.8). Social sanctions are introduced in Section 3.6. Section 3.7 concludes. We use several numerical examples to illustrate our results. In order to see the effects of the changing assumptions in the different sections most clearly, we use one example ('example 1') where parameters are the same. In other examples, parameters are chosen ad hoc in order to get nice graphs.

3.2 The model

This section describes the assumptions of the model. All propositions in this chapter are derived assuming uniformly distributed payoffs and penalties for default which are proportional to project payoffs. Throughout this chapter, we assume that project payoffs are independent. In Section 3.3, we show the first of two main results of BC. The additional assumptions we make are highlighted as Assumptions 1-3. The more specific Assumptions 4, 5 and 6 are introduced where needed, in Subsection 3.4.1, Section 3.5 and Section 3.6, respectively. Section 3.6 analyzes social sanctions, whose effects are the second main result of BC.

A given finite mass m (> 0) of risk-neutral borrowers without internal funds are endowed with one project each. The project requires an input of one unit of capital and projects are indivisible. The payoffs θ are independently and uniformly distributed: The distribution function is given by $F(\theta) = 0$ for $\theta < \underline{\theta}$, $F(\theta) = (\theta - \underline{\theta})/(\bar{\theta} - \underline{\theta})$ for $\theta \in [\underline{\theta}, \bar{\theta}]$, and $F(\theta) = 1$ for $\theta > \bar{\theta}$, where

$$\frac{\bar{\theta}}{2} > \underline{\theta} > 0.$$

The first inequality ensures a minimum degree of variation of project payoffs. The second inequality means that there is some positive payoff even in the worst case. In particular, payoffs can never be negative. As is typical of microcredit markets, we assume that MFIs offer loans without requiring collateral. At the time a loan is made, the payoff θ itself is unknown, but the distribution of payoffs is known to borrowers and banks. Once realized, the project payoff θ is common knowledge. Borrowers can neither choose between different projects nor can they influence project payoffs by the effort they spend. Thus, there is no asymmetric information: Borrowers and banks share the same information at all points in time (no hidden information).

We consider two types of contracts. With an IL contract, the borrower receives a loan of size 1 and has to make a gross repayment (principal plus loan rate) r after payoff realization.[20] A GL contract consists of a loan of size 2 to a group of two borrowers and repayment $2r$. The timing under IL and GL can be seen in Figure 3.1.

When the repayment decision is made, borrowers are endowed with sufficiently high income

[20]Note the difference in notation. In Chapter 2, r is the *net* interest rate on loans.

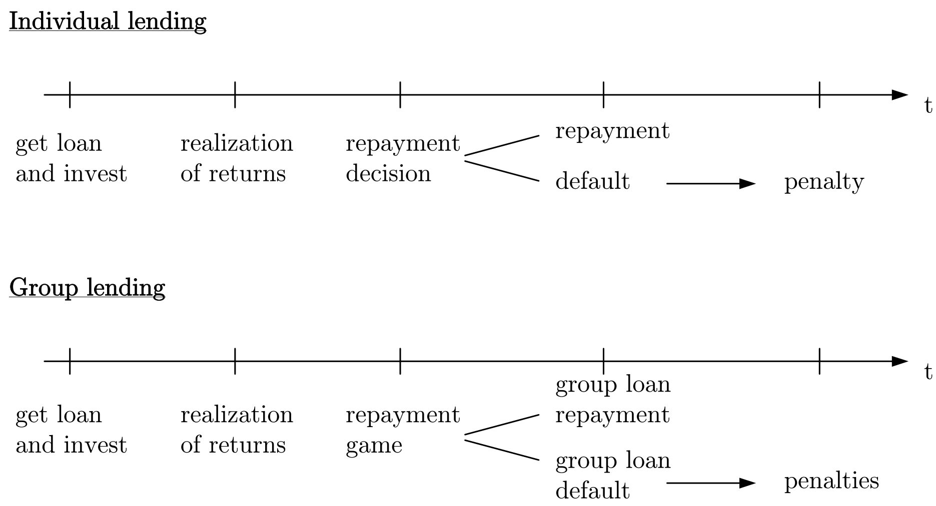


Figure 3.1: Timing of both lending schemes

so that they are able to repay.[21] We assume some exogenous second-period income a borrower could generate once returns are realized. This income can be justified by the fact that most borrowers have the possibility to sell their belongings.

However, even though borrowers are always able to repay, they do not always repay. The reason is that banks cannot perfectly enforce their claims, so that borrowers choose between repaying or not. As is common for many MFIs, repayment is an all-or-nothing decision. Thus, if a borrower with an IL contract decides to pay back, he repays r. If two borrowers with a GL contract decide to pay back, they repay $2r$ as a group or nothing at all, i.e., there are no partial repayments. If an IL borrower decides not to repay, the bank punishes him. The penalty for default is $p(\theta) = \theta/\beta$,[22] where

$$\beta > \max\{1, \underline{\theta}\}.^{23}$$

[21] Thus, we follow BC who "ignore the possibility that a borrower has insufficient funds to repay his loan" (p.4, footnote 7).

[22] There is an obvious tension between the assumptions about repayment ability on the one hand and penalties on the other hand: The penalty is independent of the value of the exogenous second-period income. As mentioned in the main text, one interpretation is that a borrower could mobilize enough money to repay by selling his belongings, but the MFI does not expect him to do this and so does not condition the penalty on the value of his belongings. An often quoted argument is the influence on banks' reputation, which makes banks reluctant to enforce such claims on the few assets of the poorest of the poor. In some countries, there are laws which prohibit seizure of basic goods.

[23] This condition captures the idea that borrowers prefer the penalty over repayment in the case of minimum project return even at zero interest: $\underline{\theta}/\beta < r = 1$ (cf. BC, p.8). Given this inequality, our former assumption

If the group defaults, the group members will be punished. It is important to note that the punishment of any one borrower does not depend on the payoff of the partner: If payoffs of two group members are θ_i and θ_j, respectively, borrower i's punishment is $p(\theta_i)$ and borrower j's is $p(\theta_j)$.

BC assume that the penalty consists of "two components", "a monetary loss due to seizure of income or assets" and "a non-pecuniary cost resulting from being 'hassled' by the bank, from loss of reputation, and so forth" (p.4). Their focus on repayment rates makes an assumption with regard to the relative magnitudes of the two components dispensable. We assume that each of the two components is a constant proportion of the penalty:

Assumption 1: *Of the penalty $p(\theta)$, a fraction α ($\in [0,1]$) is pecuniary and accrues to the MFI. The remainder of the penalty is a deadweight loss.*

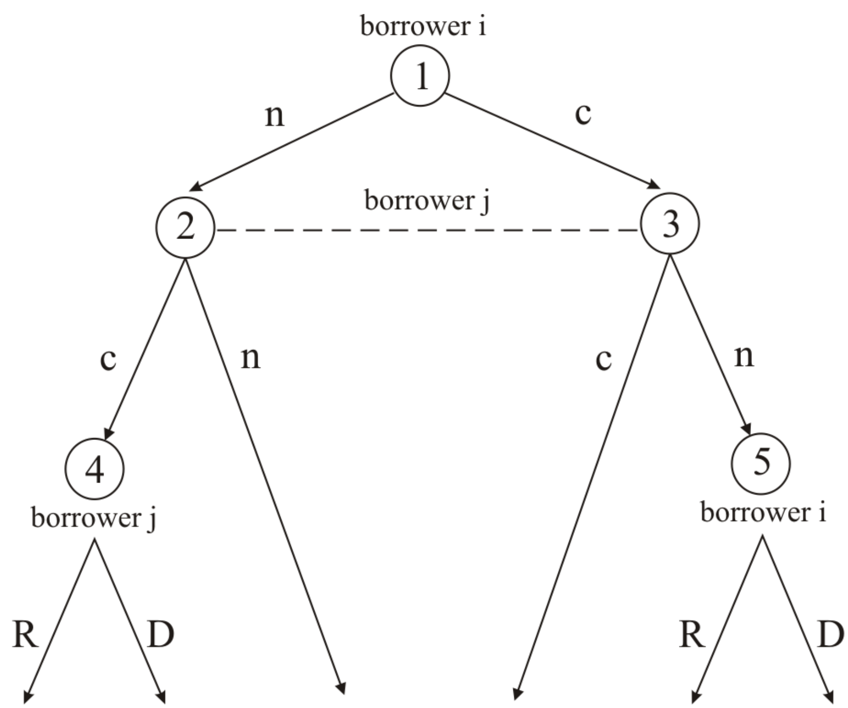
Note that α might also be negative. Exclusion from future credit as one of the most common punishments available to MFIs can cause a cost. Since, in addition, the seizure of assets to compensate for this cost is not always possible, $\alpha < 0$ is not impossible. Most of the papers on GL focus on the case $\alpha = 0$ (cf. Rai and Sjöström (2004) and Bhole and Ogden (2009)), i.e., they assume non-pecuniary penalties. Therefore, our main results are also based on this assumption. However, $\alpha > 0$ has interesting implications and might become important when the MFIs' clientele matures.[24]

The two borrowers in a group (i, j) play a two-stage repayment game. The extensive form representation can be seen in Figure 3.2. At the first stage, strategies are: contribute r to the joint repayment $2r$ (play 'c') or not (play 'n'). Both borrowers have to announce their decisions simultaneously. That is why the broken line connects nodes 2 and 3 in Figure 3.2. Together, they are an information set of borrower j.[25] When playing at the first stage, borrower j does not know what borrower i does, vice versa. If both choose to contribute, payoffs are $\theta_i - r$ and $\theta_j - r$. If both choose not to contribute, payoffs are $\theta_i - p(\theta_i)$ and $\theta_j - p(\theta_j)$, as indicated above. If borrower i chooses to contribute and j ($\neq i$) does not, i

$\bar{\theta}/2 > \underline{\theta}$ is implied by BC's (p.8) somewhat stronger condition $\bar{\theta}/2 > \beta$. Without this former assumption, the potential advantage of GL would not materialize, and IL would be unambiguously better.

[24] Using our model for a clientele with assets requires that the process of maturation is not accompanied by an elimination of the enforcement problem.

[25] We number the important nodes from 1 to 5. Information sets in the present game are, for borrower i: node 1 and node 5, for borrower j: nodes 2 and 3, node 4. Note that, since one of borrower j's information sets is not singleton, we are in a game of imperfect information.



Payoff i	θ_i	θ_i-$p(\theta_i)$	θ_i-$p(\theta_i)$	θ_i-r	θ_i-2r	θ_i-$p(\theta_i)$
Payoff j	θ_j-2r	θ_j-$p(\theta_j)$	θ_j-$p(\theta_j)$	θ_j-r	θ_j	θ_j-$p(\theta_j)$

Figure 3.2: Repayment game.

decides, at stage 2, whether he repays $2r$ alone (play 'R') or not (play 'D'). If he repays, the whole group loan is repaid and he gets $\theta_i - 2r$, whereas j gets θ_j. If not, there is group default and payoffs are $\theta_i - p(\theta_i)$ and $\theta_j - p(\theta_j)$.

We can distinguish nine strategies, which are depicted in Table 3.1.[26] This set can be searched for equilibrium strategies, contingent on the realization of payoffs.

strategy profile $(i \mid j)$	repayment?
$(c, R \mid c, R)$	yes
$(c, R \mid c, D)$	yes
$(c, D \mid c, R)$	yes
$(c, D \mid c, D)$	yes
$(n \mid c, R)$	yes
$(n \mid c, D)$	no
$(c, R \mid n)$	yes
$(c, D \mid n)$	no
$(n \mid n)$	no

Table 3.1: Strategy profiles and repayment decisions.

The supply of funds to MFIs is perfectly elastic. That is, the MFIs' cost of capital is exogenously given. One may think either of private investors with a given required rate of return or of DFIs or BVFs that can do with a below market rate of return.[27] Let q denote loan supply from MFIs to borrowers.

Assumption 2: *MFIs can raise any amount of capital $q \in [0, m]$ at the constant cost of capital ρ (≥ 1).*

The equilibrium contract maximizes expected borrower utility. In the case of funding by return-seeking investors, this is due to perfect competition among MFIs. For an MFI funded by a DFI, this is a natural objective.

[26] There is a difference between $(c, R \mid c, R)$ and $(c, D \mid c, D)$ or $(c, D \mid c, R)$. Mas-Colell, Whinston, and Green (1995, p.229) define a strategy as "a complete contingent plan that says what a player will do at each of her information sets if she is called on to play there". Thus, even though both $(c, R \mid c, R)$ and $(c, D \mid c, D)$ lead to repayment such that every borrower contributes his share, the strategies are formally different. To be fully rigorous, we would have to split up the strategies where 'n' appears and specify what a borrower would have played if he had not chosen 'n' at the first stage, i.e., there would be a difference between (n, D) and (n, R). However, playing 'n' in the first stage precludes the possibility to be called on to play in the second stage, so that we omit this distinction.

[27] Given the small proportion of microcredit markets in financial markets, an exogenous cost of capital is a natural assumption. The main results go through with an upward-sloping loan supply curve as well, and we will briefly tackle this case when we come to credit rationing in Subsection 3.4.6.

Assumption 3: *The MFIs offer the (IL or GL) contract that maximizes borrowers' expected utility subject to the constraint that it breaks even.*

3.3 Repayment rates

This section gives a brief summary of BC's results on repayment probabilities with IL and GL.

We can distinguish four qualitatively different loan rate intervals. Very low rates $r < \underline{\theta}/\beta$, low rates $\underline{\theta}/\beta \leq r \leq \bar{\theta}/(2\beta)$, high rates $\bar{\theta}/(2\beta) < r \leq \bar{\theta}/\beta$ and very high rates $\bar{\theta}/\beta < r$. However, for both IL and GL, very low and very high rates are trivial to analyze in that repayment (default) occurs with certainty for very low (high) rates. This becomes clear if we set up the condition for repayment.

Individual lending

When payoffs are realized, an IL borrower decides whether to repay or default. If the penalty which would be implied by the decision to default outweighs the amount to repay (i.e., if $p(\theta) \geq r$),[28] the borrower prefers to repay the loan. Conversely, the borrower prefers to default if the penalty is less than the contractual repayment (i.e., if $p(\theta) < r$). From the definition of the penalty function, we can also state these conditions in terms of critical payoffs. Since $p(\theta) = \theta/\beta$, an IL borrower decides to repay if, and only if, $\theta \geq \beta r$. Default occurs if, and only if, $\theta < \beta r$.

Thus, if the loan rate is very low ($r < \underline{\theta}/\beta$), an IL borrower always pays back since the penalty is higher than the loan rate for all possible payoffs (even if the worst payoff $\underline{\theta}$ is realized, the penalty is higher than r since $r < \underline{\theta}/\beta = p(\underline{\theta})$). For very high loan rates ($\bar{\theta}/\beta < r$), an IL borrower always defaults since $p(\bar{\theta}) = \bar{\theta}/\beta < r$, i.e., the penalty is lower than the loan rate even if the highest possible payoff is realized.

Therefore, we only consider loan rates $\underline{\theta}/\beta \leq r \leq \bar{\theta}/\beta$. It is convenient to define the 'default interval'

[28] Thus, we assume that repayment occurs if penalty and amount of repayment are equal.

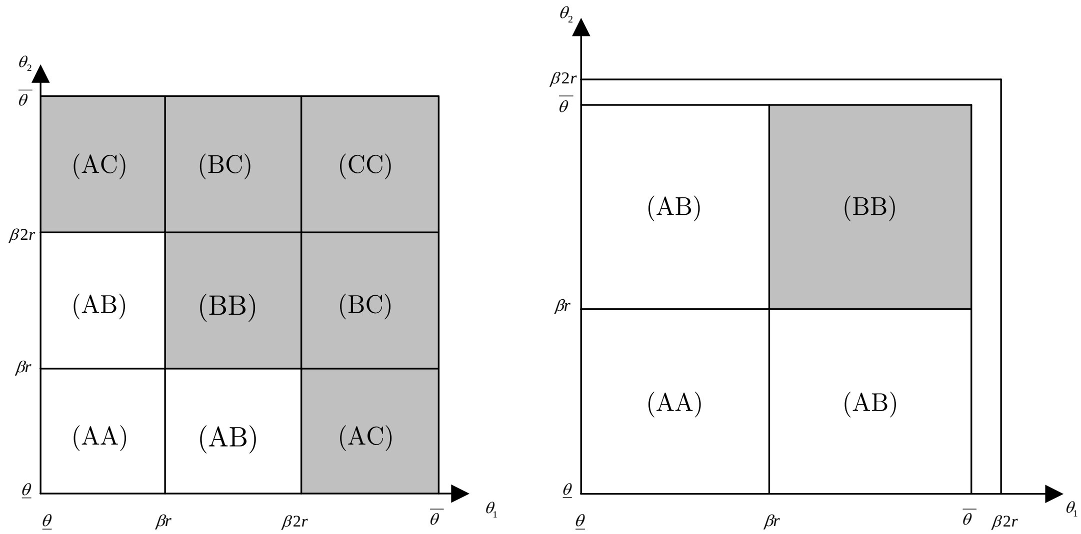


Figure 3.3: Repayment (shaded) vs. default (non-shaded), cases L (left), H (right).

$$A = [\underline{\theta}, \beta r).$$

As seen above, borrower i defaults if, and only if, $p(\theta_i) = \theta_i/\beta < r$, i.e., $\theta_i \in A$. So the repayment rate (i.e., the probability of repayment) is

$$\Pi_I(r) = 1 - F(\beta r) = \frac{\bar{\theta} - \beta r}{\bar{\theta} - \underline{\theta}}, \quad \underline{\theta}/\beta \leq r \leq \bar{\theta}/\beta. \tag{3.1}$$

This confirms the above argument about the focus on low and high loan rates in case of IL. A loan rate of $r = \bar{\theta}/\beta$ implies a zero repayment rate: $\Pi_I(\bar{\theta}/\beta) = 0$. Evidently, the repayment rate is zero for all $r > \bar{\theta}/\beta$ since $F(\beta r) = 1$ in that case. So without loss of generality, we can confine attention to loan rates $r \leq \bar{\theta}/\beta$. For $0 < r \leq \underline{\theta}/\beta$, the 'default interval' A is not well defined, the borrower repays for all θ. Such loan rates cannot arise in equilibrium, since repayment to the MFI falls short of ρ with certainty: $r \leq \underline{\theta}/\beta < 1 \leq \rho$.

Group lending

Again, we restrict attention to loan rates such that $\underline{\theta}/\beta \leq r \leq \bar{\theta}/\beta$.[29] By definition, $p(\theta_i) =$

[29] We shall see below (in footnote 39) that, as with IL, lenders' expected return falls short of ρ for loan rates $0 < r \leq \underline{\theta}/\beta$.

$\theta_i/\beta < r$ for $\theta_i \in A$, so borrower i prefers the penalty over repayment. We formally distinguish between two cases: the low-interest case

$$\text{case L:}\quad \frac{\underline{\theta}}{\beta} \leq r \leq \frac{\bar{\theta}}{2\beta},$$

and the high-interest case

$$\text{case H:}\quad \frac{\bar{\theta}}{2\beta} < r \leq \frac{\bar{\theta}}{\beta}$$

(see the left and right panel of Figure 3.3, respectively). Let

$$B = \begin{cases} [\beta r,\ 2\beta r)\,, & \text{case L,} \\ \left[\beta r,\, \bar{\theta}\right], & \text{case H.} \end{cases}$$

For $\theta_i \in B$, borrower i is willing to repay an individual loan (since $r \leq p(\theta_i) = \theta_i/\beta$) but not a group loan (since $\theta_i/\beta < 2r$). Finally, let

$$C = \left[2\beta r,\, \bar{\theta}\right], \quad \text{case L.}$$

For $\theta_i \in C$, i prefers to repay $2r$ rather than default.

BC (p.17) characterize the subgame-perfect Nash equilibria (SPNE) of the repayment game (cf. Figure 3.2 for the extensive form of the game including payoffs, and Table 3.1 for possible strategy profiles). We go through all possible payoff combinations and determine the equilibrium strategy profiles and the repayment decision entailed. To find the equilibrium, we first look for SPNE. If there is no unique SPNE, we sort out those SPNE with Pareto-inferior payoffs. If there is no Pareto dominance in payoffs, we exclude equilibria in which (at least) one borrower plays a (weakly) dominated strategy.[30]

(AA) For $(\theta_i,\, \theta_j) \in A \times A$ (cf. Figure 3.3), both borrowers have payoffs such that neither of them wants to contribute his share and $(n \mid n)$ is the only equilibrium. There is no (unilateral) incentive to deviate from $(n \mid n)$. To see this, consider one of the two borrowers (the game is

[30]The order is only important in the case with social sanctions, see below. In this section, we could also first look for equilibria in dominant strategies, and apply the concept of SPNE only in case there is no such equilibrium.

symmetric), borrower i, say. From the payoffs at the bottom of Figure 3.2, we can see that there is only one strategy profile which would make him better off, viz., $(n \mid c, R)$. In that case, his payoff would be θ_i, which is higher than $\theta_i - p(\theta_i)$. However, for borrower i to get there, borrower j would have to change his strategy. Nothing borrower i can unilaterally do leads to his preferred outcome (the same is true for borrower j), i.e., $(n \mid n)$ is a (subgame perfect)[31] Nash equilibrium. To see that it is unique, we have to check all other strategy profiles. For instance, consider $(n \mid c, R)$. Borrower j has a unilateral incentive to deviate by playing 'n' instead of 'c' at the first stage since his payoff would change from $\theta_j - 2r$ to $\theta_j - p(\theta_j)$ ($> \theta_j - 2r$ in case (AA)). The reader is invited to check that no other strategy profile is an SPNE either. The unique SPNE is $(n \mid n)$ so that the group defaults.

(BB) For $(\theta_i, \theta_j) \in B \times B$, both borrowers choosing to contribute is a Nash equilibrium and $(c, D \mid c, D)$ is subgame perfect (cf. footnote 31). Strategy pair $(n \mid n)$, i.e., both borrowers deciding not to contribute, is also an SPNE, which is however ruled out by BC (p.7) on the grounds that it is Pareto-inferior. An alternative way to get rid of this 'bad' equilibrium is elimination of weakly dominated strategies: The strategy not to contribute at the first stage is weakly dominated by the strategy to contribute.[32] Thus, case (BB) leads to repayment.

(CC) For $(\theta_i, \theta_j) \in C \times C$, which is only possible in case L, the only Nash equilibrium is that one borrower repays $2r$ and the other free-rides. This is an interesting case. Intuitively, one might guess that both borrowers contribute their shares in equilibrium. This is not the case. There are two SPNE, $(n \mid c, R)$ and $(c, R \mid n)$. Consider the wrong equilibrium guess $(c, R \mid c, R)$. Each borrower has an incentive to play 'n' at the first stage since he knows that his partner is going to play 'R' at the second stage.[33]

---

[31] Considering complete strategies in case $(n \mid n)$ for the moment, $(n, D \mid n, D)$ is subgame perfect, whereas $(n, R \mid n, R)$, $(n, R \mid n, D)$, and $(n, D \mid n, R)$ are not. If a borrower has a payoff too low to be willing to contribute his own share, he will never be willing to repay both shares, i.e., he will never play 'R' at the second stage.

[32] Since borrowers know each other's payoffs when they play the game, we use backward induction to anticipate second-stage moves before determining strategies that are dominated. Otherwise (cf. Mas-Colell, Whinston, and Green (1995, p.237)), 'c' is not weakly dominant since $(n \mid c, R)$ would yield a higher payoff, viz., θ_i.

[33] So, if both try to free-ride playing 'n' in the first stage, the group might end up defaulting even though each borrower would have preferred to repay the whole loan all by himself. However, it is not a Nash equilibrium since both borrowers would have the unilateral incentive to play (c, R). Bhole and Ogden (2009, p.4, footnote 20) explain how payment reminders from the bank after stage one rule out this kind of asymmetric equilibrium.

(BC) For $(\theta_i,\, \theta_j) \in (B \times C) \cup (C \times B)$, which is only possible in case L, the borrower i with $\theta_i \in C$ repays $2r$ and borrower j with $\theta_j \in B$ free-rides. Borrower j wants to contribute his part of the loan rather than incur the penalty, but he prefers the penalty to a repayment of $2r$. The payoff of borrower i is so high that he would rather repay the whole loan than incur the penalty. Since both know each other's payoffs, the only SPNE is $(c, R \mid n)$, borrower j free-rides on borrower i. The penalty is the worst that could happen to borrower i. Since he knows that the payoff of borrower j is insufficient to induce borrower j to repay $2r$, he will always play 'c' at the first stage. Borrower j knows that and can decide whether he wants to contribute his share r, or stay with all of his payoff. He will not contribute his part since he knows that borrower i will rather repay $2r$ than incur the penalty. The equilibrium strategy profile implies repayment.

In all these cases, the repayment received by the MFI is the same as with two IL contracts: $2r$ if the group (or both IL borrowers) repays, and $\alpha(\theta_i + \theta_j)/\beta$, the monetary part of the sum of the penalties, in case (AA).

(AB) For $(\theta_i,\, \theta_j) \in (A \times B) \cup (B \times A)$, the group defaults. This is the drawback of GL: The borrower i with $\theta_i \in B$ would repay a single loan but is discouraged from paying back anything by joint liability. In this case, BC claim that $(n \mid n)$ is the unique equilibrium. Although we concur with the claim that (AB) is a clear case of group default, there is another equilibrium strategy profile, viz., $(c, D \mid n)$.[34] Consider Figure 3.2 to verify this. Borrower j with $\theta_j \in A$ has no incentive to change his decision at stage one since he would have to contribute r instead of (the lower) $p(\theta_j)$. In comparison to payoffs when playing (c, D), borrower i would neither gain anything from playing 'n' at the first stage nor from playing 'R' at the second.[35] In any case, (AB) will lead to group default.

(AC) For $(\theta_i,\, \theta_j) \in (A \times C) \cup (C \times A)$, which is only possible in case L, the borrower i with

[34] To be more precise, $(c, D \mid n, D)$ would be the SPNE, cf. footnote 31.

[35] Note that $(n \mid n)$ is not Pareto-inferior to $(c, D \mid n, D)$. If we used elimination of (weakly) dominated strategies, $(n \mid n)$ would not even be an equilibrium. This is because (c, D) is a weakly dominant strategy (after backward induction) of borrower i. In case borrower j played 'c', borrower i would be better off, whereas he does not experience a loss if borrower j plays 'n' and he then, at the second stage, plays 'D' himself.

b_i/b_j	probability	equilibria	repayment?
(aA)	$[F(\beta r)]^2$	$(n \mid n)$	no
(aB)	$F(\beta r)[F(\beta 2r) - F(\beta r)]$	$[(n \mid n)]$ $(n \mid c, D)$	[no] no
(aC)	$F(\beta r)[1 - F(\beta 2r)]$	$(n \mid c, R)$	yes
(bA)	$[F(\beta 2r) - F(\beta r)]F(\beta r)$	$[(n \mid n)]$ $(n \mid c, D)$	[no] no
(bB)	$[F(\beta 2r) - F(\beta r)]^2$	$(c, D \mid c, D)$ $[(n \mid n)]$	yes [no]
(bC)	$[F(\beta 2r) - F(\beta r)][1 - F(\beta 2r)]$	$(n \mid c, R)$	yes
(cA)	$F(\beta r)[1 - F(\beta 2r)]$	$(c, R \mid n)$	yes
(cB)	$[1 - F(\beta 2r)][F(\beta 2r) - F(\beta r)]$	$(c, R \mid n)$	yes
(cC)	$[1 - F(\beta 2r)]^2$	$(n \mid c, R)$ $(c, R \mid n)$	yes yes

Table 3.2: Payoff combinations, probabilities, equilibrium strategies and repayment decision.

$\theta_i \in C$ repays $2r$. This is the advantage of GL: Borrower i stands in for his fellow group member. The borrower j with $\theta_j \in A$ has no incentive to contribute his share, so he will play 'n' at stage one. Borrower i prefers repaying the whole loan to defaulting and incurring his individual penalty $p(\theta_i)$. The only equilibrium is $(c, R \mid n)$ so that there is group repayment.

Table 3.2 summarizes all possible payoff combinations[36] with the respective probabilities and shows the equilibrium (or equilibria), as well as the group's repayment decision for each payoff combination. In columns three and four, we put in brackets the equilibria and repayment decisions which we can exclude using either Pareto dominance or elimination of (weakly) dominated strategies. The table is valid for case L. In case H, only cases (aA), (aB), (bA), and (bB) are possible. The corresponding probabilities in case H follow from using $F(\beta 2r) = 1$.[37]

[36] We split up (AB) into (aB) and (bA), thus not making use of the symmetry of the game for the sake of clarity in terms of the equilibrium strategy profiles in column three.

[37] In Table 3.2, note that $F(\beta 2r) = 1$ implies that the probabilities of the cases which cannot occur in case H become zero.

In case L, the repayment rate is equal to the cumulated probability of cases BB, CC, BC, and AC[38]:

$$\Pi_G(r) = 2[1 - F(2\beta r)]F(\beta r) + [1 - F(\beta r)]^2 = \frac{-3\beta^2 r^2 + 4\beta\underline{\theta} r + \bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{(\bar{\theta} - \underline{\theta})^2}, \quad \text{case L.} \tag{3.2}$$

In case H, the repayment rate is the probability of case (BB):

$$\Pi_G(r) = [1 - F(\beta r)]^2 = \left(\frac{\bar{\theta} - \beta r}{\bar{\theta} - \underline{\theta}}\right)^2, \quad \text{case H.} \tag{3.3}$$

As with IL, we can confine attention to $r \leq \bar{\theta}/\beta$, because all higher loan rates imply a zero repayment rate.[39]

The BC result

BC's (p.8) main result for the model without social sanctions is that if $\bar{\theta}/(3\beta) > 1$, then GL dominates IL in terms of repayment rates for low loan rates $r < \bar{\theta}/(3\beta)$, and vice versa. This follows from equations (3.1), (3.2), and (3.3): $\Pi_G(r) > \Pi_I(r)$ for $r \in [1, \bar{\theta}/(3\beta))$ (note that $r \leq \bar{\theta}/(3\beta)$ is case L) and $\Pi_I(r) > \Pi_G(r)$ for $r \in (\bar{\theta}/(3\beta), \bar{\theta}/\beta]$. If $\bar{\theta}/(3\beta) \leq 1$, IL yields an unambiguously higher repayment rate.

Let us note a conceptual problem comparing Π_I and Π_G: They do not measure the same thing. The former is the probability that a single borrower repays <u>one</u> unit of capital plus interest ($= r$). In contrast, the latter is the probability that a group of two borrowers repays <u>two</u> units of capital plus interest ($= 2r$). It is tempting to simply think of an IL borrower as taking a loan of size 2 such that his revenues would be in the interval $[2\underline{\theta}, 2\overline{\theta}]$. However, this would contradict one of the basic assumptions of the model: Investment opportunities are assumed to be of fixed size in that they require one unit of capital (no up- or downscaling), and that each borrower has one such project. The correct benchmark for GL is the situation

[38] Or, using the case distinctions in Table 3.2, the cumulated probability of cases bB, cC, bC, cB, aC, and cA.

[39] Loan rates $0 < r < \underline{\theta}/\beta$ cannot occur in equilibrium with GL. The interval A is not well defined in this case, so only cases (BB), (BC), and (CC) can arise. The repayment rate is unity in each of these cases, so the MFIs are unable to break even: $r < \underline{\theta}/\beta < 1 \leq \rho$.

with two independent IL borrowers, both receiving a loan of size one. In this case, it is possible that only one of the two repays his loan, whereas partial repayment is excluded for groups.

We can find the probability that both loans are repaid under two IL contracts. Due to independence of project revenues, it is simply the product of the individual repayment probabilities, i.e., Π_I^2. However, this cannot be compared to Π_G since we would neglect the probability that one IL borrower repays his loan. We could account for this using $\Pi_I^2 + 2\Pi_I(1 - \Pi_I)$ as the probability to compare to Π_G. Unfortunately, this is still a comparison of apples and oranges since we would compare the probability P('two IL borrowers pay back their loans or one of them pays back his loan') to the probability P('a group pays back the whole loan').

Therefore, in general, the comparison of BC is conceptually questionable, with one exception: As long as the penalty is completely non-pecuniary, we can derive expected repayment with GL as $\Pi_G 2r$, and with IL as $\Pi_I^2 2r + 2\Pi_I(1-\Pi_I)r + (1-\Pi_I)0 = \Pi_I 2r$. In this case, the BC comparison of Π_I and Π_G is the same as a comparison of expected repayments. We conclude that the BC proposition should rather be: If penalties are non-pecuniary, GL dominates IL in terms of expected repayment for low loan rates $r < \bar{\theta}/(3\beta)$, and vice versa.

3.4 Equilibrium

This section analyzes the equilibrium of the BC model supplemented with Assumptions 1-3. We show that an equilibrium exists, consider several interesting special cases, and highlight the allocation failures that potentially arise in equilibrium.

3.4.1 Definition of equilibrium

Given the penalty function $p(\theta) = \theta/\beta$ and $\beta > 1$, all borrowers demand loans at any loan rate, so loan demand is constant. This is because the cost of a loan (i.e., either principal plus interest repayment or penalty) is less than the payoff in every state of nature: $\min\{p(\theta), r\} = \min\{\theta/\beta, r\} < \theta$.

In order to determine borrowers' expected utility, we have to make an assumption about the probability of being the borrower who repays or the free rider in case (CC). The natural assumption is that each borrower has an equal chance of being the free rider:

Assumption 4: *The probability of being a borrower who repays when $(\theta_1, \theta_2) \in C \times C$ in case L under GL is 1/2 for each borrower.*[40]

Let $\boldsymbol{\theta}_I = \theta$ and $\boldsymbol{\theta}_G = (\theta,\ \theta')$. Denote the set of realizations of $\boldsymbol{\theta}_t$ that trigger default with lending type tL ($t \in \{\text{I, G}\}$) as D_t: $D_I = A$ and $D_G = (A \times A) \cup (A \times B) \cup (B \times A)$ (see the non-shaded areas in Figure 3.3). The binary complement of D_t (i.e., the set of realizations that trigger repayment) is denoted S_t: $S_I = [\underline{\theta},\ \bar{\theta}]\backslash D_I$ and $S_G = ([\underline{\theta},\ \bar{\theta}] \times [\underline{\theta},\ \bar{\theta}])\backslash D_G$ (see the shaded areas in Figure 3.3). $\Pi_I(r)$ and $\Pi_G(r)$ are the probabilities of $\theta \in S_I$ and $(\theta_1, \theta_2) \in S_G$, respectively. Then, using Assumption 1, the MFIs' expected repayment per dollar lent with lending type tL is

$$R_t(r) = \Pi_t(r)r + [1 - \Pi_t(r)]\alpha E\left[p(\theta)|\, \boldsymbol{\theta}_t \in D_t\right], \quad t \in \{\text{I, G}\}. \tag{3.4}$$

For IL, the formula is easy to understand. Expected repayment is simply the probability of repayment, $\Pi_I(r)$, times the amount to be repaid in that case, r, plus the probability of default times the monetary part of the expected penalty conditional on default. For GL, the formula does not immediately arise. It should consist of the probability of group repayment, $\Pi_G(r)$, times the amount to be repaid in that case, $2r$, plus the probability of group default times the monetary part of the group's expected penalty conditional on default. Formally,

$$R_G(r) = \Pi_G(r)2r + [1 - \Pi_G(r)]\alpha E\left[p(\theta_i) + p(\theta_j)|\, \boldsymbol{\theta}_G \in D_G\right]. \tag{3.5}$$

Since project revenues are independent, the conditional expectation of the sum of penalties equals the sum of the conditional expectations of penalties, i.e., $E\left[p(\theta_i) + p(\theta_j)|\, \boldsymbol{\theta}_G \in D_G\right]$ equals $E\left[p(\theta_i)|\, \boldsymbol{\theta}_G \in D_G\right] + E\left[p(\theta_j)|\, \boldsymbol{\theta}_G \in D_G\right]$. Since the penalty function is linear, the sum of the expectations of penalties can be written as the sum of the penalties of the expectations, both conditional on group default, i.e., $E\left[p(\theta_i)|\, \boldsymbol{\theta}_G \in D_G\right] + E\left[p(\theta_j)|\, \boldsymbol{\theta}_G \in D_G\right]$ equals $p(E\left[\theta_i|\, \boldsymbol{\theta}_G \in D_G\right]) + p(E\left[\theta_j|\, \boldsymbol{\theta}_G \in D_G\right])$. Moreover, θ_i and θ_j are identically distributed so that their expectations (conditional on the same event) are identical. Writing them as $E[\theta|\, \boldsymbol{\theta}_G \in D_G]$ and again using the linearity of the penalty function, the sum of the penalties of the expectations can be written as $2E\left[p(\theta)|\, \boldsymbol{\theta}_G \in D_G\right]$. Substituting in equation (3.5), the

[40] Since the number of borrowers who repay is equal to the number of borrowers who free-ride, the probability is necessarily 1/2 on average. Any mechanism that randomly assigns these roles to borrowers implies a probability of 1/2 for everyone.

latter becomes

$$R_G(r) = \Pi_G(r)2r + [1 - \Pi_G(r)]\alpha 2E\left[p(\theta)|\, \boldsymbol{\theta}_G \in D_G\right].$$

After division by two to get expected repayment per borrower, equation (3.4) obtains for $t = G$.

Using Assumption 4, expected utility of a borrower who finances his project with tL is

$$U_t(r) = \Pi_t(r)E[\theta - r|\, \boldsymbol{\theta}_t \in S_t] + [1 - \Pi_t(r)]E\left[\theta - p(\theta)|\, \boldsymbol{\theta}_t \in D_t\right], \quad t \in \{\text{I, G}\}. \tag{3.6}$$

As with expected repayments, the formula for expected utility is easy to understand for IL. Expected utility is the probability of repayment times expected net[41] revenue given repayment, plus the probability of default times expected net revenue given default. For GL, this is less clear. For a group, we should have

$$U_G(r) = \Pi_G(r)E\left[\theta_i + \theta_j - 2r|\, \boldsymbol{\theta}_G \in D_G\right] + [1 - \Pi_t(r)]E\left[\theta_i - p(\theta_i) + \theta_j - p(\theta_j)|\, \boldsymbol{\theta}_G \in D_G\right]. \tag{3.7}$$

However, similar rearrangements as with $R_G(r)$ yield expected utility per group member in equation (3.6) for $t = G$.

We have to distinguish between two types of equilibria:

Definition 3.1 *A lending type, a loan rate, and a quantity of loans* (tL, r, q) *are a* loan market equilibrium with market clearing *(also:* trade equilibrium*) if*
(1) the amount of loans made is equal to demand: $q = m$*;*
(2) MFIs make zero profit: $R_t(r) = \rho$*;*
(3) no alternative contract that attracts borrowers yields positive profit: There is no $(t'L, r') \neq (tL, r)$ *such that* $R_{t'}(r') > \rho$ *and* $U_{t'}(r') \geq U_t(r)$*.*

Definition 3.2 *A* loan market equilibrium without trade *(also:* no trade equilibrium*) prevails if there is no contract that breaks even:* $R_t(r) < \rho$ *for all* (tL, r)*.*

[41] In case of repayment, 'net revenues' are project revenues after debt redemption. In case of default, 'net revenues' refers to project revenues after punishment.

3.4.2 Existence of equilibrium

Expected repayment and expected utility can be written as functions of the loan rate for both IL and GL. This is extensive algebra based on equations (3.1),(3.2),(3.3), using the facts that project revenues are uniformly distributed and that penalties are $p(\theta) = \theta/\beta$. We do this in Appendix 3.8.1. We get

$$R_I(r) = \frac{-\beta(2-\alpha)r^2 + 2\bar{\theta}r - \frac{\alpha\underline{\theta}^2}{\beta}}{2(\bar{\theta}-\underline{\theta})}, \tag{3.8}$$

$$U_I(r) = \frac{\beta r^2 - 2\bar{\theta}r + \bar{\theta}^2 - \left(1-\frac{1}{\beta}\right)\underline{\theta}^2}{2(\bar{\theta}-\underline{\theta})}, \tag{3.9}$$

for IL, and

$$R_G(r) = \frac{-\beta^2(6-5\alpha)r^3 + 4\beta\underline{\theta}(2-\alpha)r^2 + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta} - \alpha\underline{\theta}^2)r + \frac{\alpha\underline{\theta}^3}{\beta}}{2(\bar{\theta}-\underline{\theta})^2}, \quad \text{case L}, \tag{3.10}$$

$$U_G(r) = \frac{\beta^2 r^3 - 4\beta\underline{\theta}r^2 + (2\underline{\theta}^2 - 2\bar{\theta}^2 + 4\underline{\theta}\bar{\theta})r + \bar{\theta}^3 - \underline{\theta}\bar{\theta}^2 - \underline{\theta}^2\bar{\theta} + \left(1-\frac{1}{\beta}\right)\underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2}, \quad \text{case L}, \tag{3.11}$$

and

$$R_G(r) = \frac{\beta^2(2-\alpha)r^3 - \beta\bar{\theta}(4-\alpha)r^2 + \bar{\theta}^2(2+\alpha)r - \frac{\alpha}{\beta}(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2}, \quad \text{case H}, \tag{3.12}$$

$$U_G(r) = \frac{-\beta^2 r^3 + 3\beta\bar{\theta}r^2 - 3\bar{\theta}^2 r + \bar{\theta}^3 - \left(1-\frac{1}{\beta}\right)(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2}, \quad \text{case H}, \tag{3.13}$$

for GL. Notice that $R_G(r)$ is continuous at $r = \bar{\theta}/(2\beta)$ (see Appendix 3.8.2 for the algebra). Moreover, $R_I(\underline{\theta}/\beta) = R_G(\underline{\theta}/\beta) = \underline{\theta}/\beta$ and $R_I(\bar{\theta}/\beta) = R_G(\bar{\theta}/\beta) = (\alpha/\beta)(\bar{\theta}+\underline{\theta})/2$ (see Appendix 3.8.3 for the algebra).

Let us explain the impact of an increase in α at this point. From the above formulas for expected utility, it can be seen that α does not directly affect utility. This is because the penalty function is assumed to capture all direct effects from defaulting on borrowers. However, there is an indirect effect which makes higher α a good thing. It decreases the deadweight loss from non-pecuniary penalties so that the equilibrium loan rates are lower, all other things - the cost of capital in particular - being equal. Thus, equilibrium expected utility as a function of the MFIs' cost of capital depends on α whereas expected utility as a function of the loan rate does not.

Equations (3.8)-(3.13) will be used to characterize the equilibria of the types defined in Definitions 3.1 and 3.2. To pave the way for our equilibrium analysis of the BC model, we first prove the existence of an equilibrium:

Proposition 3.1 *Either a loan market equilibrium with market clearing or a loan market equilibrium without trade exists.*

Proof: For lending type t, let r_t denote the minimum loan rate in the interval $[\underline{\theta}/\beta, \bar{\theta}/\beta]$ such that $R_t(r) = \rho$ (see Figures 3.4 and 3.8 below), and R_t^{max} be the maximum expected repayment. Since $R_t(\underline{\theta}/\beta) = \underline{\theta}/\beta < 1 \leq \rho$ and the $R_t(r)$ functions are polynomials, if $\max_{r,t} R_t(r) \geq \rho$, then r_t exists for at least one $t \in \{I, G\}$. If r_t exists for exactly one lending type tL, denote this type as t'L. If both r_I and r_G exist, let t'L be the lending type that yields higher borrower utility $U_{t'}(r_{t'})$ (if the borrower utilities are identical, pick t'L arbitrarily). We assert that $(t'\mathrm{L}, r_{t'}, m)$ is an equilibrium with market clearing. Conditions (1) and (2) in Definition 3.1 are satisfied. Clearly, if $r_{t'} = \bar{\theta}/\beta$, it is not possible to raise expected repayment beyond ρ. So consider $r_{t'} < \bar{\theta}/\beta$. By construction, $R_{t'}(\tilde{r}) > \rho$ requires $\tilde{r} > r_{t'}$. From (3.9), (3.11), and (3.13), $U_t'(r) < 0$ for all $r < \bar{\theta}/\beta$ and for $t \in \{I, G\}$ (see Appendix 3.8.4). So $U_{t'}(\tilde{r}) < U_{t'}(r_{t'})$ whenever $R_{t'}(\tilde{r}) > \rho$. That is, MFIs cannot make a positive profit with lending type t'. If r_t, $t \neq t'$, exists (i.e., if it is possible to break even with the other lending type as well), to make a profit $R_t(\bar{r}) > \rho$ with lending type tL, MFIs must set $\bar{r} > r_t$. As $U_t'(r) < 0$, this implies $U_t(\bar{r}) < U_t(r_t) \leq U_{t'}(r_{t'})$. This proves condition (3) in Definition 3.1. If $\max_{r,t} R_t(r) < \rho$, from Definition 3.2, there is a loan market equilibrium without trade. q.e.d.

Proposition 3.1 ensures that an equilibrium exists for all admissible parameter values. More importantly, the proof of the proposition is constructive: Equilibria with market clearing are found by looking for the minimum break-even loan rates for the two lending types and comparing the corresponding expected utilities of borrowers.

3.4.3 Special cases

We have already commented on the comparison of repayment rates in BC. At the end of Section 3.3, we asserted the appropriateness of the BC comparison if penalties $p(\theta)$ are completely non-pecuniary, i.e., if $\alpha = 0$. In this special case, equation (3.4) becomes $R_t(r) = \Pi_t(r)r$, so the lending type tL that yields the higher repayment rate at r also yields the higher expected repayment at r. However, even so, it is not straightforward to determine the optimal lending type. To see this, consider the following example:

Example 1: $\alpha = 0$, $\underline{\theta} = 0.6$, $\bar{\theta} = 5.5$, $\beta = 1.2$, $\rho = 1.1$. The minimum break-even loan rate is lower with GL than with IL: $r_G \approx 1.3331 < 1.4198 \approx r_I$ (see Figure 3.4). Since $r_G \approx 1.3331 < 2.2917 = \frac{\bar{\theta}}{2\beta}$, we are in case L. The repayment rate with GL (3.2) is higher than with IL (3.1): $\Pi_G(r_G) \approx 82.52\% > \Pi_I(r_I) \approx 77.47\%$. However, the associated borrower expected utilities satisfy $U_I(r_I) = 1.7338 > 1.7176 = U_G(r_G)$. Thus, the equilibrium entails IL, even though MFIs can break even with GL at a lower loan rate. Put differently, the equilibrium deadweight loss $E[\theta] - U_t(r_t) - R_t(r_t)$ caused by the non-pecuniary nature of the penalty is higher with GL (0.2324) than with IL (0.2162). We are in case L and case (AB), in which GL is disadvantageous, occurs with probability $(2F(\beta r)[F(2\beta r) - F(\beta r)] =)$ 0.1332. In that case, with GL, the expected penalty for the borrower with $\theta \in B$ is $(E[\theta|\, \theta \in B]/\beta =)$ 1.9997 – way beyond the contractual repayment. As a result, the expected penalty averaged over both borrowers is $(E[\theta|\, \boldsymbol{\theta}_G \in D_G]/\beta =)$ 1.4581 and still exceeds the loan rate $r_G = 1.3331$. This compares with an expected penalty of 0.91655 with IL. This inefficiency is a result of the group members' non-cooperative behavior in the repayment game. Another consequence is that repayment rates are rather low compared to reported rates in reality, which well exceed 90 %. We will see how this changes when we consider cooperative behavior and social sanctions.

The example proves:

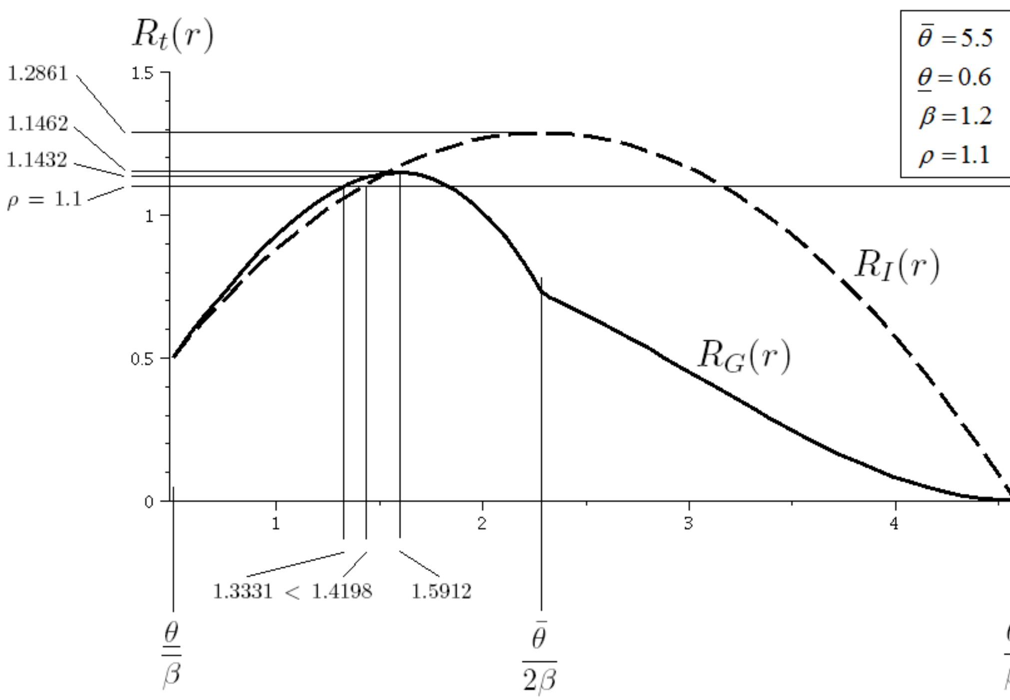


Figure 3.4: Example 1: expected repayment with non-pecuniary penalties.

Proposition 3.2 *There exist parameter values such that* (IL, r_I, m) *is a loan market equilibrium with market clearing, even though* $r_G < r_I$.

Example 1 demonstrates that the fact that GL breaks even at a lower loan rate does not mean that it occurs in equilibrium. This raises the question of whether repayment rates as an indicator of market outcomes are systematically biased in favor of GL. To answer this, we ask over which range of deposit rates ρ IL is the equilibrium mode of finance despite having the higher break-even loan rate. We address this question in two steps. First, we generalize Example 1. Then we conduct a systematic analysis of the parameter space.

Example 1 (ctd.): Let the parameters except ρ be as in Example 1. The maximum expected repayment that can be generated with GL is 1.1462 (at a loan rate of 1.5912, see Appendix 3.8.5). As can be seen from Figure 3.4, for all $\rho < 1.1432$, GL has the lower break-even loan rate. By contrast, the comparison of expected borrower utilities shows that GL occurs in equilibrium only for deposit rates up to $\rho = 1.0798$. Figure 3.5 shows this graphically. The upper panel depicts the break-even loan rate as a function of the cost of capital, ρ, for IL and GL. The lower panel shows expected utility for IL and GL at their respective break-even loan rate.

For $\rho > 1.2861$, there is a no trade equilibrium. For $1.1462 < \rho < 1.2861$ (area 'd' in Figure 3.5), only IL is possible. For $1.1432 < \rho < 1.1462$ (area 'c'), MFIs can break even with both lending types, but IL has the lower equilibrium loan rate and is the equilibrium contract. The case mentioned in Proposition 3.2 occurs in area 'b'. For $1.0798 < \rho < 1.1432$, GL leads to the lower break-even loan rate, but the equilibrium consists of IL contracts, as can be seen in the lower panel of Figure 3.5: $U_I(r_I) > U_G(r_G)$ in area 'b'. In quantitative terms, for $((1.1432 - 1.0798)/(1.1432 - 1) =)$ 44.27% of the range of ρ for which GL breaks even at a lower loan rate, IL is still the equilibrium mode of finance.

Going one step further, is the result of Proposition 3.2 an artifact of the parameters chosen in Example 1? To investigate this issue, we consider a wide array of model parameters. We pick $\underline{\theta}$, $\bar{\theta}$, and β from specified intervals, maintaining the assumption $\alpha = 0$. Consider $\underline{\theta}$-values in $[0.01, 1]$, β-values in $[1.01, 3]$, and $\bar{\theta}$-values in $[2.01, 5]$. From each of these intervals, we choose eleven values, viz.,

$$\underline{\theta} \in \{0.01, 0.1, 0.2, \ldots, 0.9, 1.0\},$$

$$\bar{\theta} \in \{2.01, 2.3, 2.6, \ldots, 4.7, 5.0\},$$

$$\beta \in \{1.01, 1.2, 1.4, \ldots, 2.8, 3.0\}.$$

We choose the values in each set to be ordered and equidistant, except the respective distance between the first and the second value (in order to always comply with the inequality assumptions on parameters dictated by the model). We build all possible triples, where each element of a triple must belong to a different one of the three sets. This gives $(11^3 =)$ 1,331 cases. For each case, we compute the interval of cost of capital $\rho \geq 1$ (if it exists) which gives rise to $r_G < r_I$ and the subinterval for which IL is nonetheless used in equilibrium. Comparing the length of the latter subinterval to the length of the former interval gives the proportion of values of cost of capital with IL in equilibrium conditional on GL having the lower break-even loan rate for the given case (i.e, the figure comparable to the 44.27% reported for Example 1). Averaging over the cases gives the mean proportion of instances in which IL occurs in

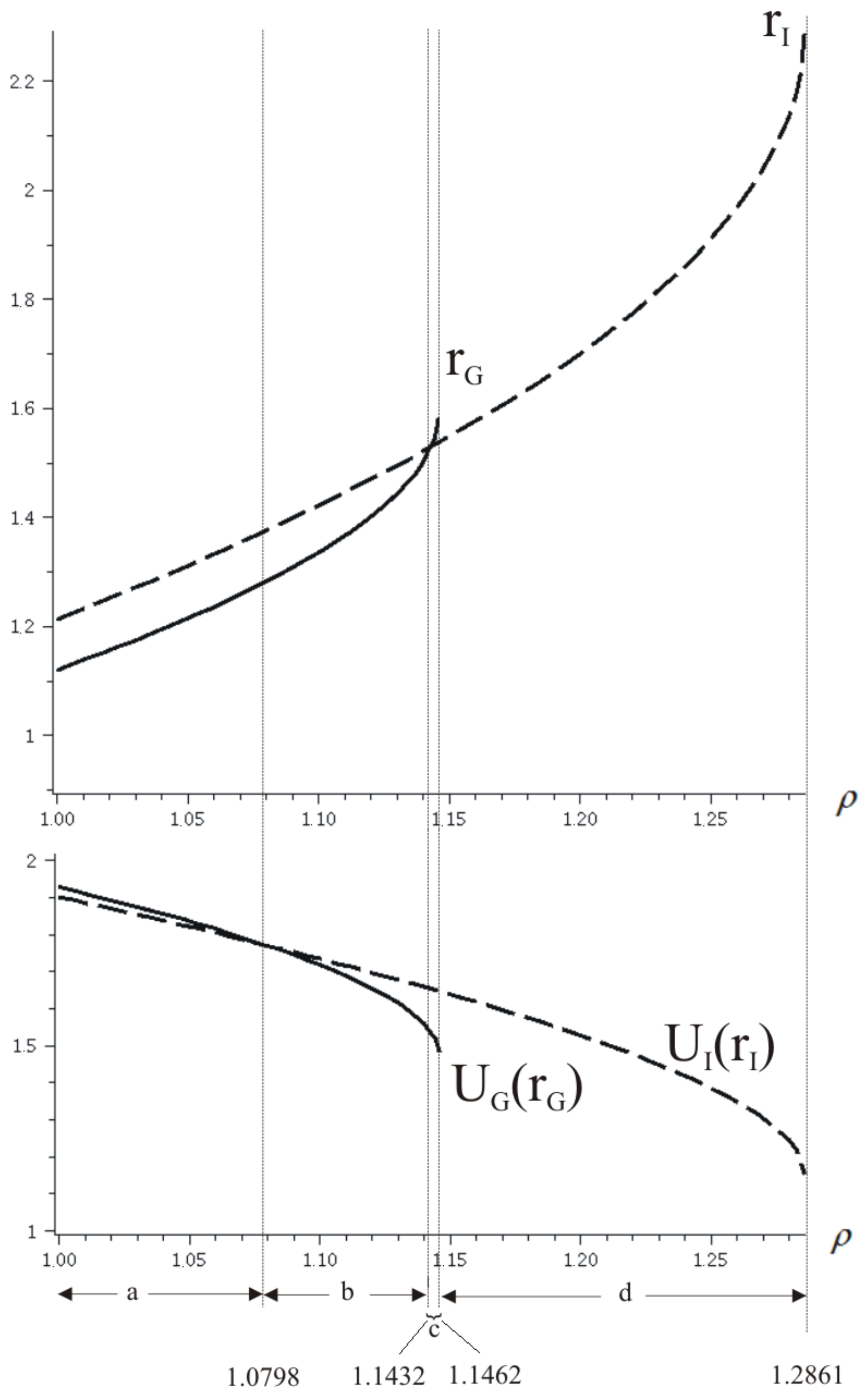


Figure 3.5: Upper panel: equilibrium loan rates, lower panel: equilibrium expected utilities.

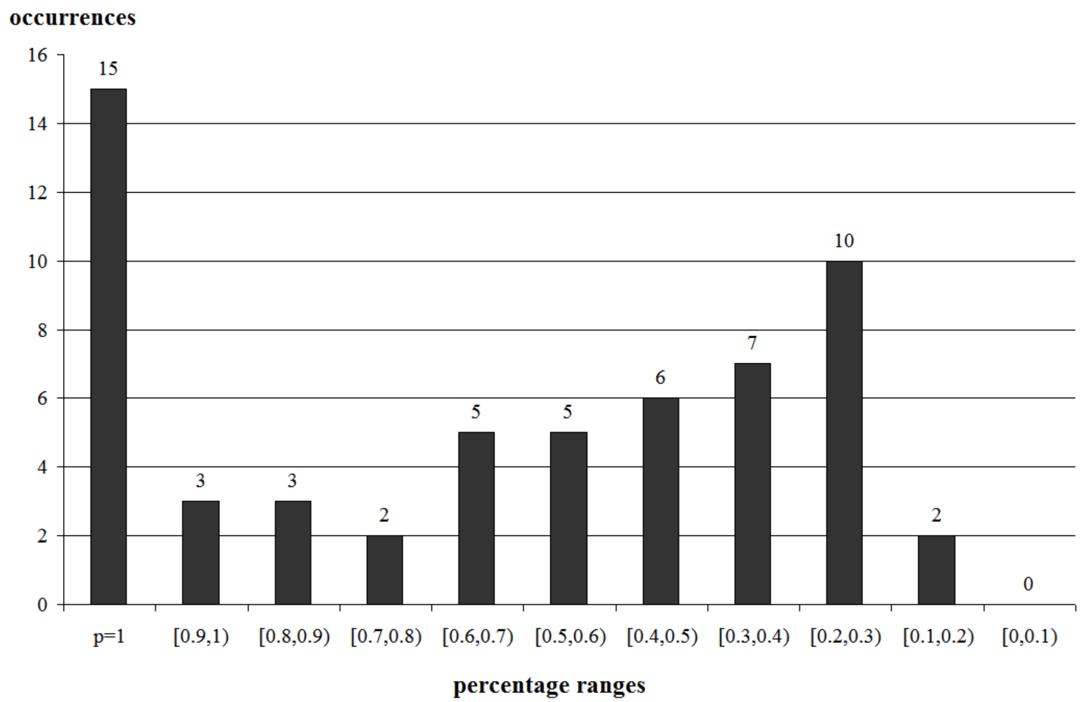


Figure 3.6: A distribution of percentages of IL equilibria in spite of $r_G < r_I$.

equilibrium despite the higher equilibrium loan rate. In $1,219$ of the $1,331$ cases, there is no trade in equilibrium. This is because projects are not good enough (low $\underline{\theta}$ and/or low $\bar{\theta}$), and/or because the enforcement problem is too strong (high β). In 47 cases, there is an equilibrium with trade for some $\rho > 1$, but there is no $\rho > 1$ such that MFIs can break even with GL contracts. In another 7 cases, there is no $\rho > 1$ such that $r_G < r_I$ and, thus,[42] the equilibrium consists of IL contracts for all $\rho \in [1, R_I^{max})$, even though MFIs can break even with GL contracts for some $\rho > 1$ (R_t^{max} is the maximum expected repayment with lending type t). The remaining 58 cases are of interest concerning Proposition 3.2. In all 58 cases, there are some $\rho > 1$ such that $r_G < r_I$. The proportion of ρ with IL in equilibrium conditional on GL having the lower break-even loan rate is shown in Table 3.3 in column four. Figure 3.6 shows a distribution of these percentages, grouping values into intervals of length 0.1, with one additional group, viz., values of 100%. The average over all 58 cases is 61.7%, even more than 44.27%.

Clearly, there is some arbitrariness in the choice of the intervals. Using $\underline{\theta} \in [0.01, 1]$,

[42]Note that $U_I(r) > U_G(r)$ in cases L and H (see Appendix 3.8.7) and that $U_t'(r) < 0$ for all $r < \bar{\theta}/\beta$ and for $t \in \{I, G\}$ (see Appendix 3.8.4).

minimum project payoff	maximum project payoff	beta	percentage
0.01	4.70	1.01	100.00
0.10	4.70	1.01	100.00
0.20	4.40	1.01	100.00
0.30	4.40	1.01	100.00
0.40	5.00	1.20	100.00
0.50	4.10	1.01	100.00
0.50	5.00	1.20	100.00
0.60	5.00	1.20	100.00
0.70	3.80	1.01	100.00
0.70	4.70	1.20	100.00
0.80	4.70	1.20	100.00
0.90	3.50	1.01	100.00
0.90	4.40	1.20	100.00
1.00	3.20	1.01	100.00
1.00	4.10	1.20	100.00
0.60	4.10	1.01	97.96
1.00	4.40	1.20	94.62
0.40	4.40	1.01	94.00
0.80	3.80	1.01	88.74
0.20	4.70	1.01	86.86
0.01	5.00	1.01	80.07
0.70	5.00	1.20	74.45
0.90	4.70	1.20	70.68
0.50	4.40	1.01	65.20
1.00	3.50	1.01	64.93
0.30	4.70	1.01	64.20
0.10	5.00	1.01	63.47
0.70	4.10	1.01	63.43
0.80	5.00	1.20	55.80
0.90	3.80	1.01	55.66
1.00	4.70	1.20	51.70
0.20	5.00	1.01	50.94
0.40	4.70	1.01	50.28
0.60	4.40	1.01	49.36
0.80	4.10	1.01	46.47
0.90	5.00	1.20	44.36
0.30	5.00	1.01	42.17
0.50	4.70	1.01	41.01
1.00	3.80	1.01	40.21
0.70	4.40	1.01	39.45
1.00	5.00	1.20	36.68
0.90	4.10	1.01	36.46
0.40	5.00	1.01	35.78
0.60	4.70	1.01	34.45
0.80	4.40	1.01	32.72
0.50	5.00	1.01	30.97
1.00	4.10	1.01	29.86
0.70	4.70	1.01	29.62
0.90	4.40	1.01	27.87
0.60	5.00	1.01	27.24
0.80	4.70	1.01	25.93
0.70	5.00	1.01	24.29
1.00	4.40	1.01	24.22
0.90	4.70	1.01	23.03
0.80	5.00	1.01	21.90
1.00	4.70	1.01	20.69
0.90	5.00	1.01	19.93
1.00	5.00	1.01	18.28

Table 3.3: Numerical results.

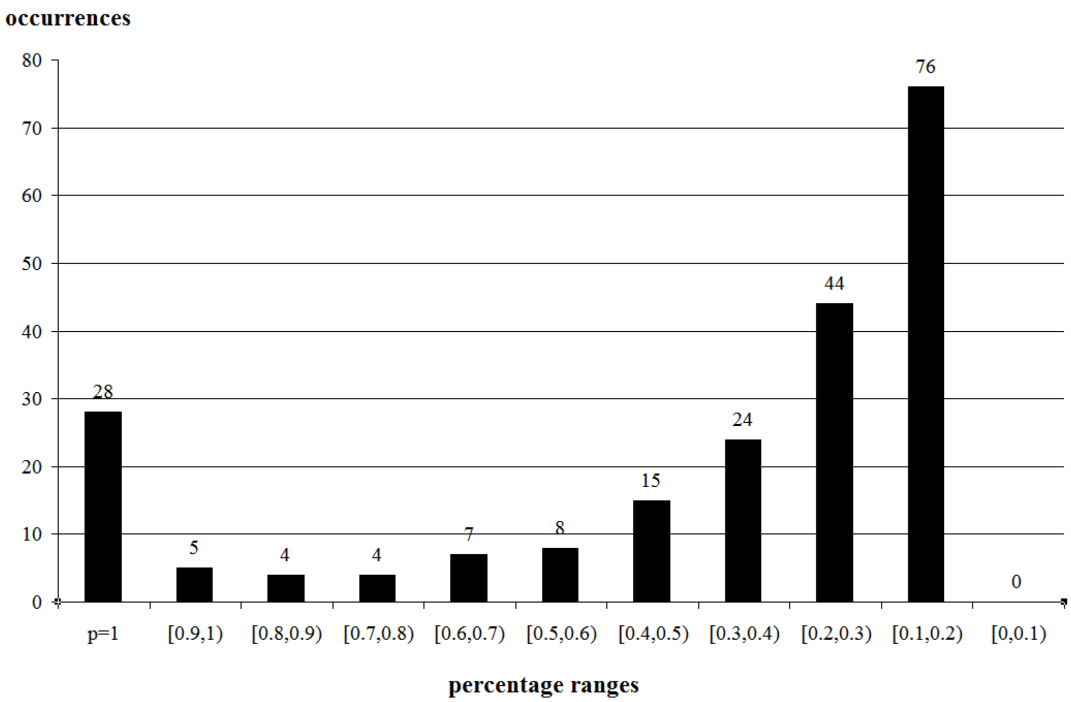


Figure 3.7: Another distribution of percentages of IL equilibria in spite of $r_G < r_I$.

$\bar{\theta} \in [2.01, 9]$, and $\beta \in [1.01, 4]$ (eleven equidistant values again), we get 1,049 cases where there is no trade in equilibrium. In 65 cases, there is trade for some $\rho > 1$, but only with IL contracts. In 2 cases, trade is possible with both lending types for some $\rho > 1$, but there is no $\rho > 1$ such that GL occurs in equilibrium. In 215 cases, we have both IL and GL equilibria for some $\rho > 1$ each. In each of these 215 cases, we can calculate our percentage of interest concerning Proposition 3.2. The distribution of these percentages can be found in Figure 3.7. The average percentage is 39.5% in this case.

By now, we have a feel on the generality of the surprising result that the equilibrium consists of IL even though GL breaks even at a lower loan rate. From this analysis, we can conclude that this phenomenon is more than just an artifact of some weird parameter constellation. However, it is striking that these cases only occur when the enforcement problem is weak (β close to one).

Another interesting special case is $\alpha = 1$. In this case, the penalties are 100 percent pecuniary, so there is no deadweight loss, and all zero-profit contracts are equally good from the

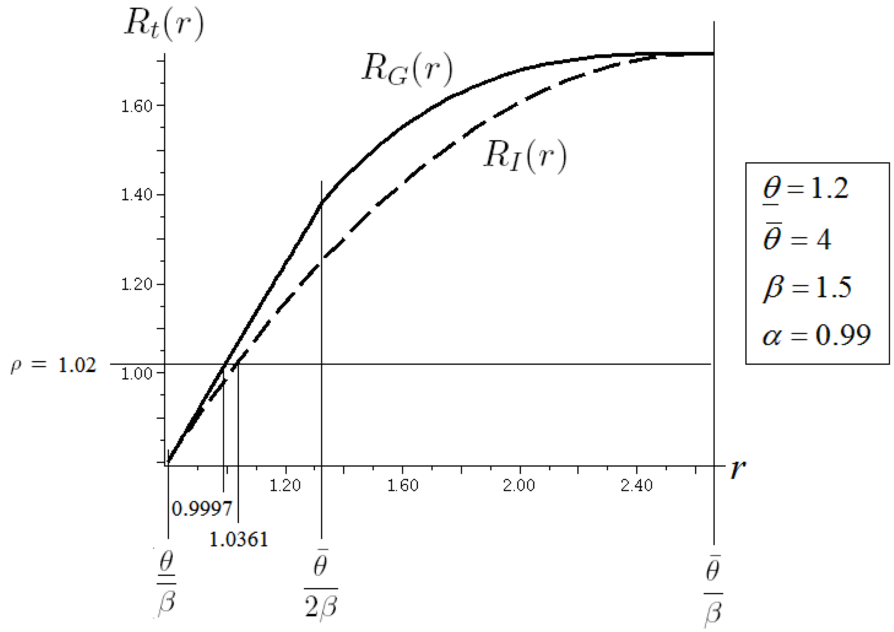


Figure 3.8: Example 2: expected repayment with pecuniary penalties ($\alpha = 0.99$).

perspective of the borrowers: $R_t(r) + U_t(r) = E[\theta]$ and $U_t(r) = E[\theta] - \rho$ when $R_t(r) = \rho$. From (3.8), (3.10), and (3.12), it follows that $R_G(r) > R_I(r)$ for all $\underline{\theta}/\beta < r < \bar{\theta}/\beta$ when $\alpha = 1$ (an algebraic proof of this statement can be found in Appendix 3.8.6). That is, the GL break-even rate r_G is lower than the IL break-even rate r_I for all ρ that give rise to loan market clearing (i.e., $\rho < (\bar{\theta} + \underline{\theta})/(2\beta)$, the maximum of both R_I and R_G if $\alpha = 1$). From the fact that $R_I(\underline{\theta}/\beta) = R_G(\underline{\theta}/\beta)$ and $R_I(\bar{\theta}/\beta) = R_G(\bar{\theta}/\beta)$ for all α and continuity of the functions on the right-hand sides of (3.8), (3.10), and (3.12) in α, it follows that for α sufficiently close to unity, if it is possible to break even with both lending types, then GL generally entails the lower break-even loan rate. One might suspect that, therefore, GL is unambiguously better than IL for α large. Interestingly, however, the assertion of Proposition 3.2 also holds true for α close to one, as the following example shows.

Example 2: $\alpha = 0.99$, $\underline{\theta} = 1.2$, $\bar{\theta} = 4$, $\beta = 1.5$, $\rho = 1.02$. The zero-profit loan rates are $r_G = 0.9997$ and $r_I = 1.0361$ (see Figure 3.8). The fact that $r_G < 1$ is interesting by itself: it shows that since (as pointed out in Example 1) the expected penalty may exceed the loan rate, a loan rate $r < 1$ may suffice to make a return $\rho > 1$. The associated expected utility levels for borrowers are $U_G = 1.5785$ and $U_I = 1.5788$.

We now turn to the allocation failures that might afflict market equilibria.

3.4.4 Financial fragility

There is 'financial fragility' as in Mankiw (1986), in that a small increase in the cost of capital potentially induces a complete breakdown of the loan market. This follows immediately from the proof of Proposition 3.1: For $\max_{r,t} R_t(r) = \rho$, an equilibrium with loan market clearing exists; as soon as ρ rises, we have $\max_{r,t} R_t(r) < \rho$, and the equilibrium entails no trade. For instance, in Example 1, $\max_{r,t} R_t(r) = 1.2861$, so the market collapses when the lenders require a return in excess of 28.61 percent. More generally, in the case $\alpha = 0$, the maximum required return beyond which the market collapses can be calculated explicitly:

Proposition 3.3 *Suppose $\alpha = 0$. Then*

$$\max_{r,t} R_t(r) = \frac{\bar{\theta}^2}{4\beta(\bar{\theta} - \underline{\theta})}. \tag{3.14}$$

When ρ rises above this value, the unique equilibrium becomes one without trade.

Proof: From (3.8) with $\alpha = 0$, $R_I(r)$ attains its maximum at $r = \bar{\theta}/(2\beta)$, and the maximum value is given by the right-hand side of (3.14). It can be shown that the functions on the right-hand sides of (3.10) and (3.12) fall short of this value in cases L and H, respectively. We do the algebra in Appendix 3.8.8. q.e.d.

The final two paragraphs of this section deal with two slight variations of the model, which give rise to redlining and rationing.

3.4.5 Redlining

Assume there is a (non-empty) finite set J of observationally distinguishable classes of borrowers of the type introduced in Section 3.2. Parameters, variables, and functions referring to class $j \in J$ are distinguished by a superscript j. For instance, $R_t^j(r)$ gives the expected repayment at a loan rate r with lending type t to type-j borrowers. Let $m = \sum_{j \in J} m^j$, so that Assumption 2 states that the supply of capital at ρ is sufficient to finance all projects of all classes. An equilibrium prevails if for each type $j \in J$, the conditions of either Definition

3.1 or Definition 3.2 are satisfied. *Redlining* is said to prevail when there is an equilibrium with market clearing for some types j and a no trade equilibrium for others.

Proposition 3.4 *Redlining prevails if, and only if,*

$$\min_j \max_{r,t} R_t^j(r) < \rho \leq \max_j \max_{r,t} R_t^j(r).$$

Proof: The equilibrium lending type, loan rate, and loan volume for each class are found following the same steps as in the proof of Proposition 3.1. Types j with $\rho \leq \max_{r,t} R_t^j(r)$ get a loan via IL or GL, and at an equilibrium loan rate specific to that class of borrowers. Types j with $\max_{r,t} R_t^j(r) < \rho$ do not. So the condition in Proposition 3.4 ensures that some classes get loans, while others do not. q.e.d.

The most interesting case arises when classes differ not with regard to payoffs but with regard to the nature and magnitude of the penalties: $\underline{\theta}^j = \underline{\theta}$ and $\bar{\theta}^j = \bar{\theta}$ for all j, but the α^j's and/or β^j's differ. In this case, if the condition of the proposition is satisfied, some borrowers do not get credit, even though others with equally good projects do.

Example 3: There are three classes: $J = \{1, 2, 3\}$. Penalties are non-pecuniary, the cost of capital and the payoff boundaries are: $\rho = 1.02$, $\alpha^j = 0$, $\bar{\theta}^j = 6$, and $\underline{\theta}^j = 1$ for all j. The penalty parameters are $\beta^1 = 1.25$, $\beta^2 = 1.5$, and $\beta^3 = 2$. Class-1 borrowers get loans with GL at $r_G^1 = 1.0613$. Borrowers of type $j = 2$ get individual loans at $r_I^2 = 1.2254$. For class 3, there is no way to break even: $\max_{r,t} R_t^3(r) = 0.9$. Due to the limited scope for punishment after non-repayment, borrowers in this class are redlined.

The model with observationally distinguishable borrower classes has an important implication for the roles of for-profit organizations and of DFIs which can do with lower returns: To maximize the amount of credit given, DFIs have to direct their funds under management to banks that finance classes with maximum expected repayment just above the MFIs' required gross return, while private investors target the high-yield market segments, for otherwise the DFIs would crowd out private investment.

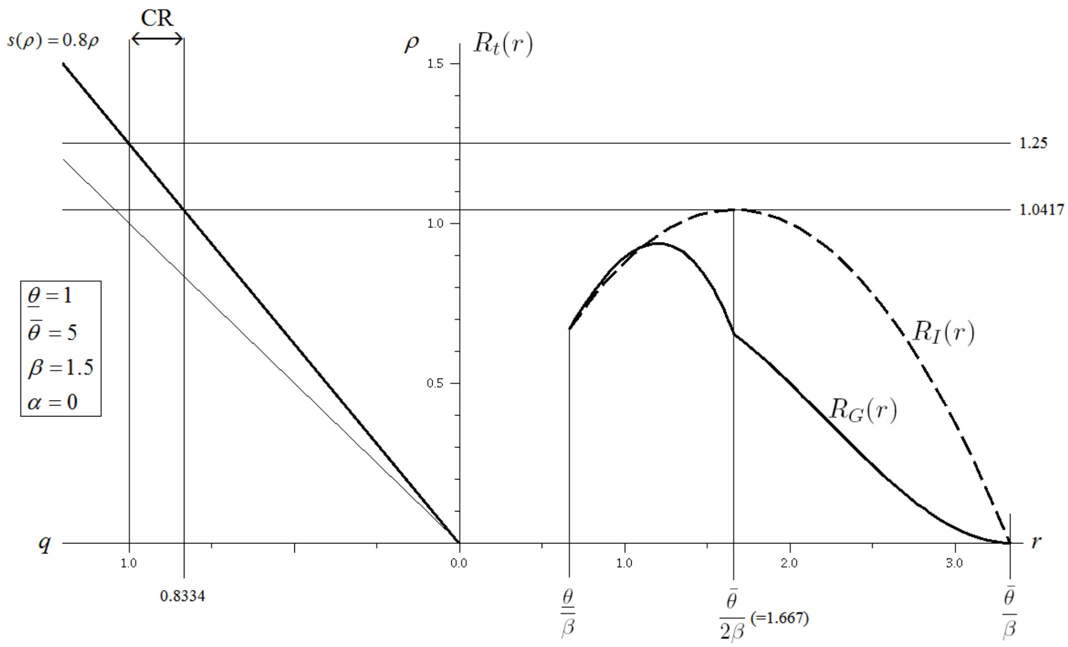


Figure 3.9: Example 4: credit rationing.

3.4.6 Credit rationing

In Chapter 2, we have seen how excess demand for credit can be an equilibrium phenomenon. By considering capital supply with uncertain interest, we have shown that a globally hump-shaped return function is not necessary for credit rationing. In the model of this chapter, there is neither project dependence nor capital risk. However, we have seen in Figure 3.4 that the return function can be hump-shaped. We now show how credit rationing can arise.

Going back to the one-class case, assume now, instead of Assumption 2, that the loan supply is a real-valued, strictly increasing function $s(\rho)$. A lending type, a loan rate, and a quantity of loans $(t\mathrm{L}, r, q)$ are a *credit rationing equilibrium* if (1) $0 < q < m$, (2) and (3) in Definition 3.1 are satisfied, and (4) $q = s(\rho)$. A credit rationing equilibrium occurs when the supply of funds at the loan rate which maximizes expected repayment falls short of the demand for credit:

Proposition 3.5 *If $s\left[\max_{r,t} R_t(r)\right] < m$, a credit rationing equilibrium occurs.*

Example 4: Let $\alpha = 0$, $\underline{\theta} = 1$, $\bar{\theta} = 5$, and $\beta = 1.5$. Further, let $m = 1$ and $s(\rho) = 0.8\rho$. The maximum expected repayment $\max_{r,t} R_t(r) = 1.0417$ is obtained with IL at $r_I = 1.6667$. The corresponding loan supply is $s(1.0417) = 0.8334$. So in equilibrium, MFIs make individual loans to 83.34 percent of the borrowers at the loan rate $r_I = 1.6667$ (see Figure 3.9). The projects' expected revenue (i.e., $(\bar{\theta} + \underline{\theta})/2 = 3$) is far beyond the level needed to induce savers to supply enough capital to finance all projects (viz., $\rho = 1.25$), but the enforcement problem leads to credit rationing. In Figure 3.9, the panel on the right shows the familiar graph of expected repayment as a function of the loan rate. In the panel on the left, capital supply is depicted as a function of ρ, where the independent variable, ρ, is on the ordinate. The thin decreasing (or, from right to left: increasing) line in the left panel is the 45-degree line, i.e., capital supply if we had $s(\rho) = \rho$. Since the abscissa on the left is a function of the ordinate, $s(\rho) = 0.8\rho$, the thick decreasing (increasing) line, must be steeper than the 45-degree line. In equilibrium, MFIs make zero profits so that $\max_{r,t} R_t(r)$ equals ρ. Drawing a horizontal line from maximum expected repayment into the left panel yields an intersection with capital supply. The value of the intersection on the abscissa gives us equilibrium capital supply. Since credit demand is normalized to one, the difference $(1 - 0.83 =)$ 0.17 is the amount of credit rationing ('CR' in Figure 3.9).[43] MFIs have no incentive to increase the loan rate since this would decrease repayment rates such that expected repayments to the MFIs would be lower.

3.5 Cooperative behavior

The analysis so far lends support to the widely held view that the scope for GL is rather limited if the sole characteristic of GL is joint liability. In Section 3.1, we have explained the important concept of side trades. Up to now, we have only considered contracts between a bank and (one or two) borrowers. In this section, we follow Ahlin and Townsend (2007, Subsection 1.3.2) and assume that borrowers play the repayment game cooperatively rather than non-cooperatively.

3.5.1 Repayment game and expected repayments

Assumption 5: *The borrowers 1 and 2 in a group repay iff* $2r \leq p(\theta_1) + p(\theta_2)$. *They share the net payoffs* $\theta_1 + \theta_2 - \min\{2r, p(\theta_1) + p(\theta_2)\}$ *such that both have the same expected utility.*

[43] In Chapter 2, we have seen the distinction of credit rationing types by Keeton (1979). As in Chapter 2, the assumption of indivisibility of projects implies that rationing is of type II.

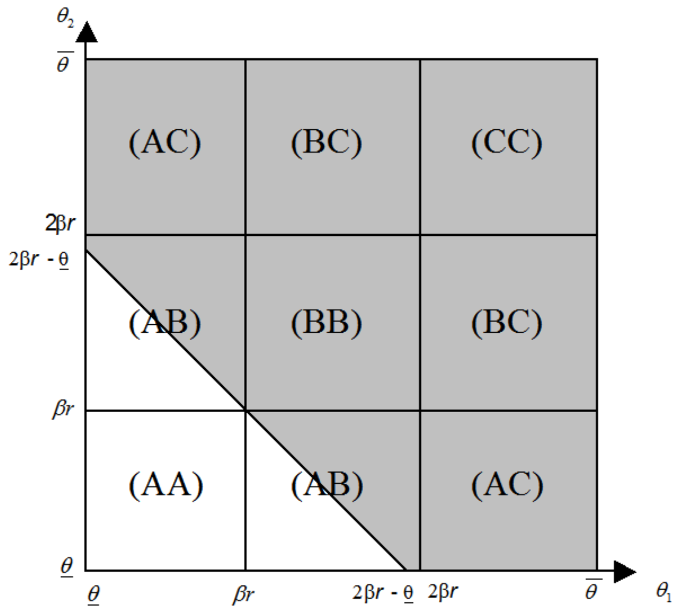


Figure 3.10: Repayment (shaded) vs. default (non-shaded) with cooperative behavior

Using $p(\theta) = \theta/\beta$, Assumption 5 implies that the members of a group repay whenever $\theta_1+\theta_2 \geq 2\beta r$. Equal expected utilities could be achieved, for example, by sharing $\theta_1+\theta_2-\min\{2r, (\theta_1+\theta_2)/\beta\}$ equally for all possible realizations (θ_1, θ_2).

In terms of Figure 3.10, the two members of a group repay for (θ_1, θ_2) on and above the line with slope -1 through $(\beta r, \beta r)$. As in the non-cooperative repayment game, the members of a group repay nothing in case (AA) and $2r$ in cases (BB), (BC), (CC) and (AC). Cooperation is conducive to the repayment rate under GL because in the fields corresponding to case (AB) in Figure 3.10, they default only in the area below the line $\theta_1 + \theta_2 = 2\beta r$. One can infer from Figure 3.10 that the expected repayment with GL is higher than with IL: The repayments with GL or IL are the same in cases (AA), (BB), (BC), and (CC). With GL, both group members repay in field (AB) above the line $\theta_1 + \theta_2 = 2\beta r$ and in field (AC), whereas both default in field (AB) below the line $\theta_1 + \theta_2 = 2\beta r$. With IL, by contrast, one of two borrowers repays in cases (AB) and (AC). Since the area of the non-shaded 'default triangle' (the non-shaded area less (AA)) with GL is less than half the total area of fields (AB) and (AC), given the uniform distribution of (θ_1, θ_2), the expected repayment is higher with GL than with IL (cf. Ahlin and Townsend, 2007, Proposition 8, p.F24). This is true as long as

$\frac{\theta}{\beta} < r < \frac{\bar{\theta}+\theta}{2\beta}$.

We focus on the case of non-pecuniary penalties, i.e., $\alpha = 0$. Thus, the repayment rate is $1 - [(2\beta r - \underline{\theta}) - \underline{\theta}]^2/[2(\bar{\theta} - \underline{\theta})^2]$, and the expected repayment is

$$R_C(r) = \frac{(\bar{\theta} - \underline{\theta})^2 - 2(\beta r - \underline{\theta})^2}{(\bar{\theta} - \underline{\theta})^2} r \tag{3.15}$$

for

$$\frac{\underline{\theta}}{\beta} \leq r \leq \frac{\bar{\theta} + \underline{\theta}}{2\beta}. \tag{3.16}$$

$R_C(r)$ has the familiar hump shape (see Appendix 3.8.9). The loan rate that maximizes the expected repayment is denoted as r_C^{max} and the corresponding expected repayment as R_C^{max} ($\equiv R_C(r_C^{max})$). In the main text, we restrict attention to loan rates that satisfy (3.16) so that we avoid the distinction between cases L and H. In Appendix 3.8.10 we show that the analysis readily extends to loan rates $r > (\bar{\theta} + \underline{\theta})/(2\beta)$. Notice that $\frac{\underline{\theta}}{\beta} \leq r \leq \frac{\bar{\theta}+\underline{\theta}}{2\beta}$ implies $\frac{\underline{\theta}}{\beta} \leq r \leq \frac{\bar{\theta}}{2\beta}$ since $(\bar{\theta} + \underline{\theta})/(2\beta) < \bar{\theta}/\beta$. From (3.8) and (3.15), $R_C(r) \geq R_I(r)$ for all r such that (3.16) holds (see Appendix 3.8.9). Let r_C be the minimum break-even loan rate with GL (provided that $R_C^{max} \geq \rho$). The fact that $R_C(r) \geq R_I(r)$ implies $r_C \leq r_I$ whenever break-even with IL is possible (i.e., whenever $R_I^{max} \geq \rho$).

3.5.2 Equilibrium

Proposition 3.6 *When borrowers act cooperatively in the repayment game, (GL, r_C, m) is the unique equilibrium whenever $R_C^{max} \geq \rho$.*

Proof: As in case of non-cooperative behavior in the repayment game, borrowers demand loans at any loan rate. With either IL or GL, they could always default and be better off than without having done the project(s). Thus, $q = m$ whenever there is a loan rate that allows MFIs to break even. In what follows, we distinguish between Part I and Part II (cf. the upper and lower panel of Figure 3.11, respectively).

Part I: $\bar{\theta} < 7.2749 \cdot \underline{\theta}$

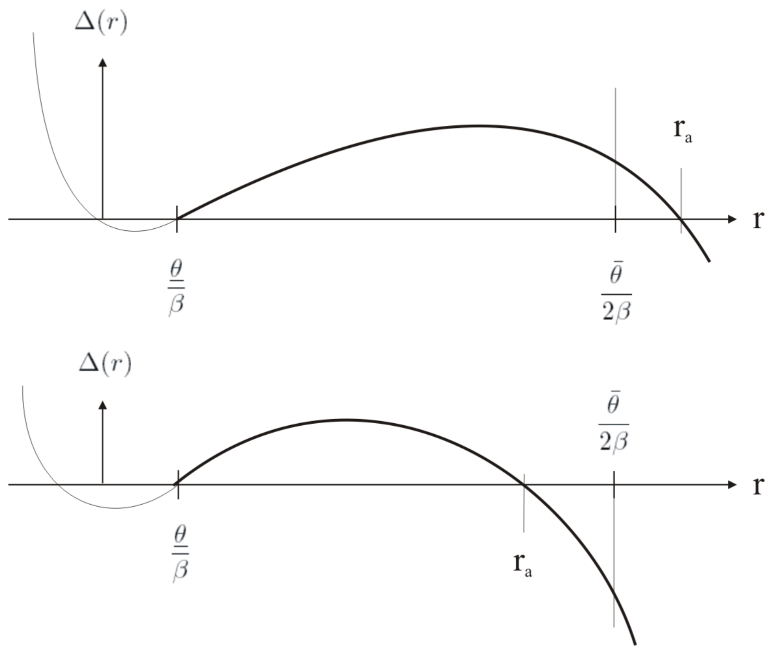


Figure 3.11: Deadweight loss difference with cooperative behavior.

For ρ such that $R_C^{max} \geq \rho > R_I^{max}$, GL is used in equilibrium because IL does not break even (and the loan rate is r_C as will become clear below). So we can focus on the case $R_I^{max} \geq \rho$. We derive the deadweight loss with GL, $D_C(r)$, for $\frac{\underline{\theta}}{\beta} < r < \frac{\bar{\theta}+\underline{\theta}}{2\beta}$. From Figure 3.10, it can be seen to be

$$2D_C(r) = \int_{\underline{\theta}}^{2\beta r-\underline{\theta}} \frac{1}{\bar{\theta}-\underline{\theta}} \int_{\underline{\theta}}^{2\beta r-\theta_1} \frac{\frac{\theta_1}{\beta}+\frac{\theta_2}{\beta}}{\bar{\theta}-\underline{\theta}} d\theta_2 d\theta_1 = \frac{4\beta^3 r^3 - 6\beta^2 \underline{\theta} r^2 + 2\underline{\theta}^3}{3\beta(\bar{\theta}-\underline{\theta})^2}, \qquad (3.17)$$

so that we have

$$D_C'(r) = \frac{4\beta(\beta r - \underline{\theta})r}{(\bar{\theta}-\underline{\theta})^2} > 0$$

for $r > \underline{\theta}/\beta$. $D_C(r)$ satisfies

$$E[\theta] - D_C(r) = U_C(r) + R_C(r). \qquad (3.18)$$

For $r = r_C$, this becomes $E[\theta] - D_C(r_C) = U_C(r_C) + \rho$. To achieve $R_C(\tilde{r}) > \rho$ and $U_C(\tilde{r}) \geq U_C(r_C)$ with GL at $\tilde{r} \neq r_C$, the deadweight loss must be smaller, $D_C(\tilde{r}) < D_C(r_C)$. However, since r_C is the minimum break-even loan rate, $R_C(\tilde{r}) > \rho$ requires $\tilde{r} > r_C$. Since $D'_C(r) > 0$ for all $r > \underline{\theta}/\beta$, we have $D_C(\tilde{r}) > D_C(r_C)$, a contradiction. So there is no profitable GL contract that attracts borrowers (which also proofs the use of r_C if banks can only break even with GL, see above).

Similarly, an IL contract with $R_I(\bar{r}) > \rho$ and $U_I(\bar{r}) \geq U_C(r_C)$ must satisfy $D_I(\bar{r}) < D_C(r_C)$ and, as r_I is the minimum break-even loan rate, $\bar{r} > r_I$. We derive a contradiction. Without loss of generality, we can assume $r \leq \bar{\theta}/(2\beta)$. This is because for any loan rate above $\bar{\theta}/(2\beta)$ that breaks even, there is a lower loan rate that yields the same expected repayment but a lower deadweight loss. The latter follows from

$$D_I(r) = \frac{\beta^2 r^2 - \underline{\theta}^2}{2\beta(\bar{\theta} - \underline{\theta})}, \tag{3.19}$$

which implies that $D'_I(r) > 0$ for all $r > 0$. Using $\bar{r} > r_I \geq r_C$ and $D'_I(r) > 0$, $D_I(r) \geq D_C(r)$ is sufficient to prove that there is no profitable IL contract, since it implies that

$$D_I(\bar{r}) > D_I(r_I) \geq D_I(r_C) \geq D_C(r_C),$$

a contradiction. So it remains to show that $D_C(r) \leq D_I(r)$ for all $r \leq \bar{\theta}/(2\beta)$. Let $\Delta(r) \equiv D_I(r) - D_C(r)$. We have to show that $\Delta(r) \geq 0$ for $r \leq \bar{\theta}/(2\beta)$. Substituting for $D_I(r)$ and $D_C(r)$ from (3.19) and (3.17), respectively, gives

$$\Delta(r) = \frac{-8\beta^3 r^3 + 3\beta^2(\bar{\theta} + 3\underline{\theta})r^2 - \underline{\theta}^2(3\bar{\theta} + \underline{\theta})}{6\beta(\bar{\theta} - \underline{\theta})^2}.$$

The polynomial on the right-hand side has a local minimum at $r = 0$, a root at $r = \underline{\theta}/\beta$, and a local maximum at $r = (\bar{\theta} + 3\underline{\theta})/(4\beta)$. So

$$\Delta\left(\frac{\bar{\theta}}{2\beta}\right) = \underline{\theta}^3 \frac{-\left(\frac{\bar{\theta}}{\underline{\theta}}\right)^3 + 9\left(\frac{\bar{\theta}}{\underline{\theta}}\right)^2 - 4\left(3\frac{\bar{\theta}}{\underline{\theta}} + 1\right)}{24\beta(\bar{\theta} - \underline{\theta})^2} \geq 0$$

is sufficient for $\Delta(r) \geq 0$ for r in the interval $[\underline{\theta}/\beta, \bar{\theta}/(2\beta)]$. The third-order polynomial in $\bar{\theta}/\underline{\theta}$ in the numerator on the right-hand side has roots -0.2749, 2, and 7.2749. So the conditions $\bar{\theta}/2 > \underline{\theta}$ and $\bar{\theta}/7.2749 < \underline{\theta}$ ensure $\Delta(r) \geq 0$.

Part II: $\bar{\theta} > 7.2749 \cdot \underline{\theta}$

$\Delta(r)$ has the shape indicated in the lower panel of Figure 3.11. In particular, there is a loan rate $r_a < \frac{\bar{\theta}}{2\beta}$ such that $\Delta(r_a) = \Delta(\frac{\underline{\theta}}{\beta})$. The sufficient condition from Part I, $D_I(r) \geq D_C(r)$ for **all** $r < \frac{\bar{\theta}}{2\beta}$, is not necessary. It suffices to have $D_I(r_C) > D_C(r_C)$. So, in terms of the lower panel of Figure 3.11, if $\frac{\underline{\theta}}{\beta} < r_C < r_I < r_a < \frac{\bar{\theta}}{2\beta}$, and even if $\frac{\underline{\theta}}{\beta} < r_C < r_a < r_I < \frac{\bar{\theta}}{2\beta}$, we have $D_I(r_I) > D_I(r_C) > D_C(r_C)$, so that the chain of arguments of Part I applies (recall that $D_I'(r) > 0$ and $r_C < r_I$). However, if $\frac{\underline{\theta}}{\beta} < r_a < r_C < r_I < \frac{\bar{\theta}}{2\beta}$, we have to resort to another proof.

The idea is simple but the algebra is tedious: We show that $\frac{\underline{\theta}}{\beta} < r_a < r_C < r_I < \frac{\bar{\theta}}{2\beta}$ contradicts the assumption $\bar{\theta} > 7.2749 \cdot \underline{\theta}$. Suppose $\frac{\underline{\theta}}{\beta} < r_a < r_C < r_I < \frac{\bar{\theta}}{2\beta}$ holds. Interestingly, the loan rate r_a can be obtained as the larger root of $\frac{\Delta(r)}{(r-\frac{\underline{\theta}}{\beta})}$. This is

$$r_a = \frac{1}{16\beta}\left(3\bar{\theta} + \underline{\theta} + \sqrt{33\underline{\theta}^2 + 102\bar{\theta}\underline{\theta} + 9\bar{\theta}^2}\right).$$

We have

$$512\beta(\bar{\theta} - \underline{\theta})^2\left(R_C(r_a) - R_I^{max}\right)$$

$$= \sqrt{33\underline{\theta}^2 + 102\bar{\theta}\underline{\theta} + 9\bar{\theta}^2}\left(23\bar{\theta}^2 - 46\bar{\theta}\underline{\theta} - 25\underline{\theta}^2\right) - 131\bar{\theta}^2\underline{\theta} - 59\bar{\theta}^3 + 551\bar{\theta}\underline{\theta}^2 + 215\underline{\theta}^3.$$

Using $\bar{\theta} = x\underline{\theta}$, we do not change the sign if we divide by $\underline{\theta}^3$ to get the RHS as an expression in x only, viz.,

$$\phi(x) \equiv \frac{1}{(x-1)^2}\left(\sqrt{33 + 102x + 9x^2}(23x^2 - 46x - 25) - 131x^2 - 59x^3 + 551x + 215\right).$$

This function is strictly increasing for $x > 1$ and has a root at $x \approx 3.01$ (cf. the solid lines in Figure 3.12).[44] Since $R_C(r_a) - R_I^{max} > 0$ for $\bar{\theta} > 3.01 \cdot \underline{\theta}$ and, thus, for $\bar{\theta} > 7.2749 \cdot \underline{\theta}$, we have a contradiction:

$$\rho = R_C(r_C) > R_C(r_a) > R_I^{max},$$

[44] The dashed line in Figure 3.12 is the slant asymptote of $\phi(x)$. Its algebraic form is $142 + 10x$.

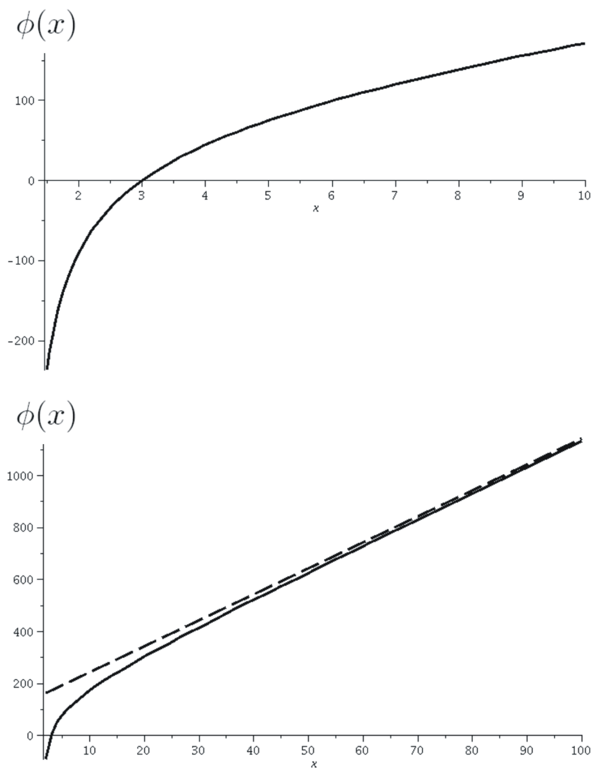


Figure 3.12: Upper panel: $\phi(x)$, lower panel: $\phi(x)$ and asymptote (dashed).

i.e., r_I does not exist.[45] q.e.d.

The proof goes through without modification when we allow for $r > (\bar{\theta}+\underline{\theta})/(2\beta)$ (see Appendix 3.8.10). So with cooperative behavior, GL not only yields the higher repayment rate but also becomes the equilibrium lending type. This result lends support to the view that other mechanisms besides joint liability are needed to fully exploit the potential of group lending.

The market equilibrium is potentially characterized by the market failures discussed in the preceding section. This follows immediately from the fact that the expected repayment function $R_C(r)$ has the familiar hump shape, so that borrowers cease to get funds when ρ rises beyond R_C^{max}, and positive excess demand at r_C^{max} does not lead to an increase in the loan rate (when the supply of capital is imperfectly elastic). Thus, GL with cooperative repayment behavior does not *eliminate* the market failures introduced in the preceding section. However, since $R_C^{max} > R_I^{max}$, the level of the cost of capital at which lending breaks down

[45] $R_C(r_C) > R_C(r_a)$ holds since we supposed $\frac{\underline{\theta}}{\beta} < r_a < r_C < r_I < \frac{\bar{\theta}}{2\beta}$ and since $R_C'(r) > 0$ for $\frac{\underline{\theta}}{\beta} < r < r_C^{max}$, and $r_C < r_C^{max}$ by the definition of r_C.

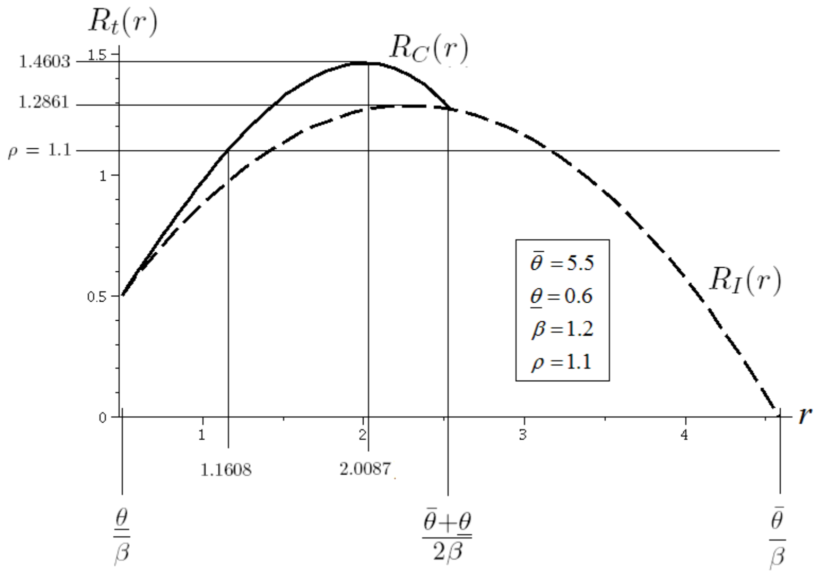


Figure 3.13: Example: IL versus GL with cooperative behavior

and the supply of capital in a rationing equilibrium are higher with GL (given positively and imperfectly elastic supply). In this sense, GL *ameliorates* the market failures.

Example 1 (ctd.): With $\underline{\theta} = 0.6$, $\bar{\theta} = 5.5$, $\beta = 1.2$, and $\rho = 1.1$, the equilibrium loan rate is $r_C = 1.1608$. While the repayment rate and expected utility with GL were equal to 82.52% and 1.7338 with non-cooperative behavior (cf. example 1), they rise to 94.76% and 1.9007, respectively. $R_C(r)$ achieves its maximum $R_C^{max} = 1.4603$ at $r_C^{max} = 2.0087$. While with non-cooperative behavior the market breaks down when ρ rises beyond 1.2861, an equilibrium exists for costs of capital up to 1.4603 here (see Figure 3.13).

3.6 Social sanctions

Following BC (Section 4), we introduce social sanctions to the model. This is motivated by the fact that borrowers in a joint liability group might affect each other's payoffs. For instance, if borrower 'Jane' wants to repay under an IL contract, but the decision of her group member 'John' discourages her from repaying anything (so that she has to incur a penalty), she is worse off than with the IL contract. BC discuss three forms of social sanctions, all based on strong social ties between community members. First, Jane might directly admonish John if John's repayment decision negatively affects her payoff. Second, since borrowers in close-knit communities usually interact in areas distinct from lending groups, too, Jane might change her behavior toward John in general. Third, Jane might tell others in that community that John behaved selfishly at her expense, so that John's reputation suffers.

The extent of social sanctions depends on several parameters of the model. BC distinguish between two main factors: the extent of harm done to the borrower who suffers and the reasonableness of the decision not to contribute. In particular, they assume that the decision not to contribute does not entail social sanctions if there is no harm done to the partner. Furthermore, there are no social sanctions for a borrower with the worst possible payoff $\underline{\theta}$. BC assume that social sanctions are an increasing function of both the harm done to the contributing borrower and the payoff of the borrower who does not contribute.

BC's main result in this regard is that if social sanctions are severe enough, GL yields a higher repayment rate than IL (BC, Proposition 3, p.12). We adopt a simple specification of social sanctions and show that, if penalties from the bank are non-pecuniary and social sanctions are strong enough, GL has a higher repayment rate than IL and that GL is the unique mode of finance. However, the loan market equilibrium still displays the allocation problems analyzed in Subsections 3.4.4-3.4.6.

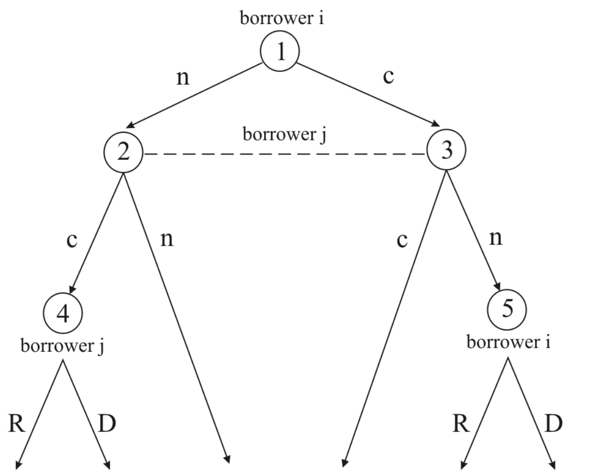
3.6.1 Repayment game and expected repayments

To analyze the effects of social sanctions, we go back to the case of non-cooperative behavior in the repayment game.

Assumption 6: *If a borrower i in a group decides to contribute at stage 1 of the repayment game, then if his fellow group member j decides not to contribute, i imposes a sanction $s > r$ on him. No sanctions are imposed otherwise.*

That is, a social sanction is imposed when one borrower's decision not to contribute forces his fellow group member to choose between repaying the group loan alone or accepting the penalty despite his declared willingness to repay his part of the loan.[46] As for the severity

[46]The assumption that no sanctions are imposed otherwise is immaterial. Adding sanctions in other instances as well strengthens our conclusions. For the sake of clarity of exposition, we choose just the minimal set of



Payoff i	θ_i-s	θ_i-p(θ_i)-s	θ_i-p(θ_i)	θ_i-r	θ_i-2r	θ_i-p(θ_i)
Payoff j	θ_j-2r	θ_j-p(θ_j)	θ_j-p(θ_j)	θ_j-r	θ_j-s	θ_j-p(θ_j)-s

Figure 3.14: Repayment game with social sanctions.

of the sanctions, since, as before, we can focus on loan rates $r \leq \bar{\theta}/\beta$, we could alternatively assume $s > \bar{\theta}/\beta$, so that the sanction simply has to be 'sufficiently large' relative to model parameters, irrespective of its specific dependence on r, θ_i, and θ_j.[47]

In contrast to part ii) of Assumption 1 in BC (p.10), an implication of our assumption about social sanctions is that a borrower has to incur social sanctions even if his payoff is the worst possible. This makes sense since we explicitly assume that borrowers are always able to repay, if not from project payoffs then from exogenous second-period income.

The presence of social sanctions strengthens the incentives to contribute in the repayment game. Figure 3.14 shows the modified repayment game in a group with social sanctions. It is similar to Figure 3.2, only the payoffs differ for some strategy profiles. Setting $s > r$ (Assumption 6), the payoffs at the bottom of Figure 3.14 become as in Table 3.4. We use the same steps as in Section 3.3 to find the equilibrium (1. SPNE, 2. Pareto dominance, 3. Elimination of weakly dominated strategies). For the cases defined in Section 3.3, the following equilibria arise:

(AA) There are two SPNE, viz., $(n \mid n)$ and $(c, D \mid c, D)$. This is interesting since neither

sanctions that make GL become the dominant mode of finance.

[47] In particular, the sanction may or may not differ depending on whether i repays $2r$ or accepts $p(\theta_i)$ at stage 2.

$(i \mid j)$	$(n \mid c, R)$	$(n \mid c, D)$	$(n \mid n)$	$(c \mid c)$	$(c, R \mid n)$	$(c, D \mid n)$
i	$< (\theta_i - r)$	$< (\theta_i - p(\theta_i) - r)$	$\theta_i - p(\theta_i)$	$\theta_i - r$	$\theta_i - 2r$	$\theta_i - p(\theta_i)$
j	$\theta_j - 2r$	$\theta_j - p(\theta_j)$	$\theta_j - p(\theta_j)$	$\theta_j - r$	$< (\theta_j - r)$	$< (\theta_j - p(\theta_j) - r)$

Table 3.4: Strategy profiles $(i \mid j)$ and payoffs of the game with social sanctions.

borrower would repay an individual loan. Without social sanctions, $(c, D \mid c, D)$ is not an equilibrium since either borrower prefers to play 'n', incurring the (small) penalty instead of the contribution r. However, a deviation from the strategy 'c' at stage 1 would imply penalty *and* sanction s when there are social sanctions. Since there are multiple equilibria, we apply our second criterion, Pareto dominance in payoffs. Comparing payoffs in columns 4 and 5 in Table 3.4, we can exclude strategy profile $(c, D \mid c, D)$ from the set of SPNE, so that the equilibrium entails group default.[48]

(AB) This is the critical case for GL. With social sanctions, there are two SPNE, viz., $(n \mid n)$ and $(c, D \mid c, D)$. However, in contrast to case (AA), we cannot rank these SPNE using the Pareto dominance criterion since the borrower i with $\theta_i \in A$ is better off with strategy profile $(n \mid n)$, whereas the borrower j with $\theta_j \in B$ prefers strategy profile $(c, D \mid c, D)$. Thus, we use elimination of dominated strategies to find a unique equilibrium. Since playing 'n' at stage 1 involves the risk of sanctions, contribution at stage 1 is weakly dominant (after backward induction, cf. footnote 32) for both borrowers. Thus, we rule out $(n \mid n)$ and are left with an equilibrium which entails group repayment.

(AC) As without social sanctions, the unique SPNE entails group repayment. However, whereas the borrower i with $\theta_i \in A$ was free-riding without social sanctions, the unique equilibrium with social sanctions is $(c, D \mid c, R)$ so that both borrowers contribute their share.

(BB) There are two SPNE, viz., $(n \mid n)$ and $(c, D \mid c, D)$. However, $(n \mid n)$ is Pareto-inferior

[48]Note the similarity to the famous prisoners' dilemma. Due to the nature of social sanctions introduced in Assumption 6, it is a weakly dominant strategy for both borrowers to contribute at stage 1. This is because playing 'n' at stage 1 involves the risk of being sanctioned. Thus, $(c, D \mid c, D)$ is an equilibrium in weakly dominant strategies, even though it is Pareto-inferior to $(n \mid n)$. This is the above-mentioned case where the order in which we use refinement criteria plays a role. In essence, our chosen order has one crucial effect, namely to solve the coordination problem in the prisoners' dilemma. It reflects the fact that two borrowers are not like two prisoners kept imprisoned in separate cells. Note that requirements for coordination differ between our situation and the prisoners' dilemma. In our case, coordination only needs to achieve a Pareto-superior Nash equilibrium, whereas the Pareto-superior strategy profile in the prisoners' dilemma is not a Nash equilibrium.

(and (c, D) is also a weakly dominant strategy for each borrower). Thus, the group repays.

(BC) As in case (AC), social sanctions avoid free-riding. The unique SPNE $(c, D \mid c, R)$ entails group repayment.

(CC) As in cases (AC) and (BC), social sanctions avoid free-riding. The unique SPNE $(c, R \mid c, R)$ entails group repayment.

Assumption 6 thus eliminates the drawback of GL: Borrowers repay unless case (AA) occurs.

The repayment rate becomes

$$\Pi_S(r) = 1 - F(\beta r)^2 = \frac{-\beta^2 r^2 + 2\beta\underline{\theta}r + \bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{(\bar{\theta} - \underline{\theta})^2} \tag{3.20}$$

(there is no need to distinguish between cases L and H). GL dominates IL in that it brings about group repayment in cases (AC) and (BC), where one of the two individual benchmark borrowers defaults. Accordingly, from (3.1) and (3.20), $\Pi_S(r) > \Pi_I(r)$ whenever $F(\beta r) < 1$, i.e., $r < \bar{\theta}/\beta$.[49] However, we want to remind the reader of the conceptual problems comparing individual and group repayment rates, cf. Section 3.3. Thus, the result itself is less interesting than its implications for a comparison of expected repayments.

As before, we can confine attention to $r \leq \bar{\theta}/\beta$ because $\Pi_t(r) = 0$ for $r > \bar{\theta}/\beta$ $(t \in \{I, S\})$. First, let us restrict attention to the case $\alpha = 0$, so that penalties are completely non-pecuniary and $R_S(r) \equiv \Pi_S(r)r > \Pi_I(r)r = R_I(r)$ for $r < \bar{\theta}/\beta$. The function $R_S(r)$ has the characteristic hump shape over the interval $(\underline{\theta}/\beta, \bar{\theta}/\beta)$ (see Appendix 3.8.11).

3.6.2 Equilibrium

In order to determine the credit market equilibrium, we need to talk about credit demand. Recall that borrowers demand loans at any loan rate when there are no social sanctions. Given non-cooperative behavior, the cost of a loan (i.e., either principal plus interest repayment or penalty) is less than the payoff in every state of nature: $\min\{p(\theta), r\} = \min\{\theta/\beta, r\} < \theta$. In case of cooperative behavior, groups can always decide to default so that imperfect

[49] $1 - F(\beta r) < 1 - [F(\beta r)]^2 \Leftrightarrow F(\beta r) < 1 \Leftrightarrow \beta r - \underline{\theta} < \bar{\theta} - \underline{\theta} \Leftrightarrow r < \bar{\theta}/\beta$.

enforcement implies that expected utility of any borrower is positive at all loan rates.

Interestingly, this is not true when there are social sanctions. Social sanctions might force borrowers to use exogenous second-period income to contribute their share of the group loan if $\underline{\theta}$ is very low. For instance, let $\underline{\theta}$ be close to zero. If $\theta_i = \underline{\theta} \in A$ and $\theta_j \in B$ are realized (case (AB)), both borrowers contribute r in equilibrium. For most loan rates, borrower i is not able to contribute r using his project payoff alone. However, instead of incurring social sanctions, he will use second-period income to contribute his share. For now, we continue to assume that borrowers are risk-neutral, but not without mentioning a critical feature of this assumption. Contrary to the case where borrowers do neither cooperate nor sanction each other, borrowers might actually use second-period income. Thus, assuming risk neutrality means that the loss of (all of) their belongings bothers borrowers as much as earning the same amount of money. This is not necessarily true and we comment on the theoretical effects of 'loss aversion' further below. With risk neutrality, borrowers demand loans if their expected utility from doing so is positive at the stipulated loan rate.

Letting r_S denote the minimum loan rate that allows MFIs to break even, and R_S^{max} the maximum expected repayment with GL and social sanctions, we can show the following

Proposition 3.7 *Let $\alpha = 0$. If $R_S^{max} \geq \rho$, the unique equilibrium is (GL, r_S, m). Otherwise the unique equilibrium entails no trade.*

Proof: With social sanctions and $\alpha = 0$, expected borrower utility with lending type t is

$$\begin{aligned} U_t(r) &= \Pi_t(r)E[\theta - r|\, \boldsymbol{\theta}_t \in S_t] + (1-\Pi_t(r))E\left[\theta - \frac{\theta}{\beta}\middle|\, \boldsymbol{\theta}_t \in D_t\right] \\ &= E[\theta] - \Pi_t(r)r - \frac{1}{\beta}(1-\Pi_t(r))E[\theta|\, \boldsymbol{\theta}_t \in D_t]. \end{aligned}$$

With $E[\theta] = (\bar{\theta}+\underline{\theta})/2$, $R_S(r) = \Pi_S(r)r$ using equation (3.20), and $E[\theta|\, \boldsymbol{\theta}_G \in D_G] = (\beta r+\underline{\theta})/2$, we get

$$U_S(r) = \frac{1}{2\beta(\bar{\theta}-\underline{\theta})^2}\left\{\beta^3 r^3 - 3\underline{\theta}\beta^2 r^2 + \beta r(-2\bar{\theta}^2 + 4\bar{\theta}\underline{\theta} + \underline{\theta}^2) - \underline{\theta}^3 + \beta[(\bar{\theta}^2 - \underline{\theta}^2)(\bar{\theta} - \underline{\theta})]\right\}. \tag{3.21}$$

The derivative w.r.t. r is

$$U'_S(r) = \frac{1}{2\beta(\bar{\theta}-\underline{\theta})^2}\left[3\beta^3 r^2 - 6\underline{\theta}\beta^2 r + \beta(-2\bar{\theta}^2 + 4\bar{\theta}\underline{\theta} + \underline{\theta}^2)\right]. \tag{3.22}$$

Setting this derivative equal to zero, we get two roots

$$r_1 = \frac{1}{2\beta}\left(2\underline{\theta} - \sqrt{\frac{8}{3}}(\bar{\theta}-\underline{\theta})\right), \quad r_2 = \frac{1}{2\beta}\left(2\underline{\theta} + \sqrt{\frac{8}{3}}(\bar{\theta}-\underline{\theta})\right), \tag{3.23}$$

where $r_1 < \frac{\underline{\theta}}{\beta}$ and $\frac{\underline{\theta}}{\beta} < r_2 < \frac{\bar{\theta}}{\beta}$. Moreover, $U_S(\frac{\underline{\theta}}{\beta}) > 0$ and $U_S(\frac{\bar{\theta}}{\beta}) > 0$. Thus, a GL borrower's expected utility has one of the two shapes indicated in the two panels of Figure 3.15. It is interesting to see that, first, expected utility is **not** decreasing all over the interval $(\frac{\underline{\theta}}{\beta}, \frac{\bar{\theta}}{\beta})$, so that borrowers might prefer higher loan rates. Second, regarding the lower panel, expected utility might be negative for some loan rates. It can be shown that there are parameter constellations where expected utility is indeed negative for some loan rate.[50] However, we show that loan rates leading to negative expected utility can never occur in equilibrium - not because there is no demand (note the tautology), but because MFIs would never offer such loan rates.

In Figure 3.15, r_b is defined as the loan rate $r > \frac{\underline{\theta}}{\beta}$ such that $U_S(r_b) = U_S(\frac{\bar{\theta}}{\beta})$. The structure of the remainder of the proof is this: First, we show that the equilibrium loan rate with GL can never exceed r_S^{max}. Second, starting from r_S, we show that there is no other GL contract that attracts borrowers and yields profits for banks. Third, we show that there is no such IL contract either. Fourth, we show that borrower demand is m for all possible $r < r_S^{max}$.

1. Since $R_S(r)$ is hump-shaped with $\alpha = 0$ (see Appendix 3.8.11), $R_S(r_S^{max}) > R_S(r)$ for all $r \in (r_S^{max}, \frac{\bar{\theta}}{\beta}]$. We now show that $U_S(r_S^{max}) > U_S(r)$ for all $r \in (r_S^{max}, \frac{\bar{\theta}}{\beta}]$ so that the equilibrium loan rate cannot be larger than r_S^{max}. If it were, there would be a contract which borrowers prefer and with profits for banks, viz., (GL, r_S^{max}).

r_S^{max} is the larger root of $R'_S(r)$:

[50] We comment on this phenomenon further below.

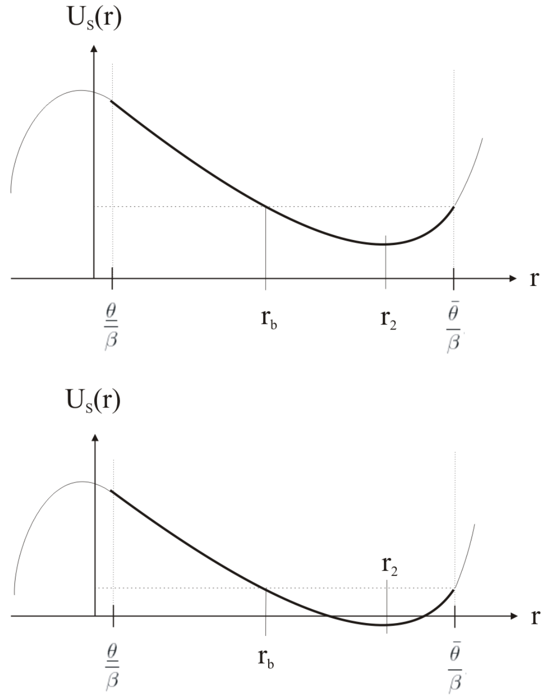


Figure 3.15: Expected utility under GL with social sanctions.

$$r_S^{max} = \frac{1}{3\beta}\left(2\underline{\theta} + \sqrt{4\underline{\theta}^2 + 3\bar{\theta}^2 - 6\bar{\theta}\underline{\theta}}\right). \tag{3.24}$$

It seems impossible to show $U_S(r_S^{max}) > U_S(\frac{\bar{\theta}}{\beta})$ which would be sufficient due to the shape of $U_S(r)$. Instead, we show that another sufficient condition for $U_S(r_S^{max}) > U_S(r)$ for all $r \in (r_S^{max}, \frac{\bar{\theta}}{\beta}]$ holds, viz., $r_S^{max} < r_b$.[51] Interestingly, we get an algebraic expression for r_b by solving $[U_S(r) - U_S(\frac{\bar{\theta}}{\beta})]/(r - \frac{\bar{\theta}}{\beta})$ for its larger root, since $U_S(r) - U_S(\frac{\bar{\theta}}{\beta})$ has a root at $\frac{\bar{\theta}}{\beta}$. We get

$$r_b = \frac{1}{2\beta}\left(-\bar{\theta} + 3\underline{\theta} + (\bar{\theta} - \underline{\theta})\sqrt{5}\right). \tag{3.25}$$

We have $r_S^{max} < r_b$ if and only if

$$16\underline{\theta}^2 + 12\bar{\theta}^2 - 24\bar{\theta}\underline{\theta} < \left[\bar{\theta}(3\sqrt{5} - 3) + \underline{\theta}(5 - 3\sqrt{5})\right]^2. \tag{3.26}$$

This inequality holds for all $\bar{\theta} > \underline{\theta}$, and, thus, since $\bar{\theta} > 2\underline{\theta}$, for all admissible parameter values.

[51] If $r_S^{max} < r_b$, $U_S(r_S^{max}) > U_S(r_b) \equiv U_S(\frac{\bar{\theta}}{\beta})$ since $U_S'(r) < 0$ for $r < r_b < r_2$.

2. Due to 1., we only have to look at $r < r_S^{max}$. Since $U_S'(r) < 0$ and $R_S'(r) > 0$ for these r, loan rates $r < r_S$ do not allow banks to break even, and loan rates $r > r_S$ do not attract borrowers.

3. From all IL contracts, r_I, if it exists, is the only possible equilibrium contract since $U_I'(r) < 0$ for all $r < \frac{\bar{\theta}}{\beta}$ (cf. the proof of Proposition 3.1). Since $R_S(r) > R_I(r)$ for all $r < \frac{\bar{\theta}}{\beta}$, we have $r_I > r_S$. The deadweight loss with GL is

$$D_S(r) = \frac{\beta^3 r^3 - \beta^2 \underline{\theta} r^2 - \beta \underline{\theta}^2 r + \underline{\theta}^3}{2\beta(\bar{\theta} - \underline{\theta})^2}, \tag{3.27}$$

so that

$$D_S'(r) = \frac{3\beta^3 r^2 - 2\beta^2 \underline{\theta} r - \beta \underline{\theta}^2}{2\beta(\bar{\theta} - \underline{\theta})^2}, \tag{3.28}$$

which is positive for all $r > \underline{\theta}/\beta$. Furthermore, since

$$D_I(r) = \frac{\beta^2 r^2 - \underline{\theta}^2}{2\beta(\bar{\theta} - \underline{\theta})}, \tag{3.29}$$

we have $D_S(r) < D_I(r)$ for all r.[52] Therefore,

$$D_S(r_S) < D_S(r_I) < D_I(r_I).$$

Using $R_S(r_S) = R_I(r_I) = \rho$, we have $D_t(r_t) = E[\theta] - \rho - U_t(r_t)$ for $r = \{S, I\}$ and, thus

$$\begin{aligned} D_S(r_S) &< D_I(r_I) \\ E[\theta] - \rho - U_S(r_S) &< E[\theta] - \rho - U_I(r_I) \\ U_S(r_S) &> U_I(r_I). \end{aligned}$$

Thus, starting from r_S, there is no profitable IL contract that attracts borrowers.

[52] Note that the RHS of equation (3.27) can be written as $\frac{(\beta r - \underline{\theta})(\beta^2 r^2 - \underline{\theta}^2)}{2\beta(\bar{\theta} - \underline{\theta})^2}$.

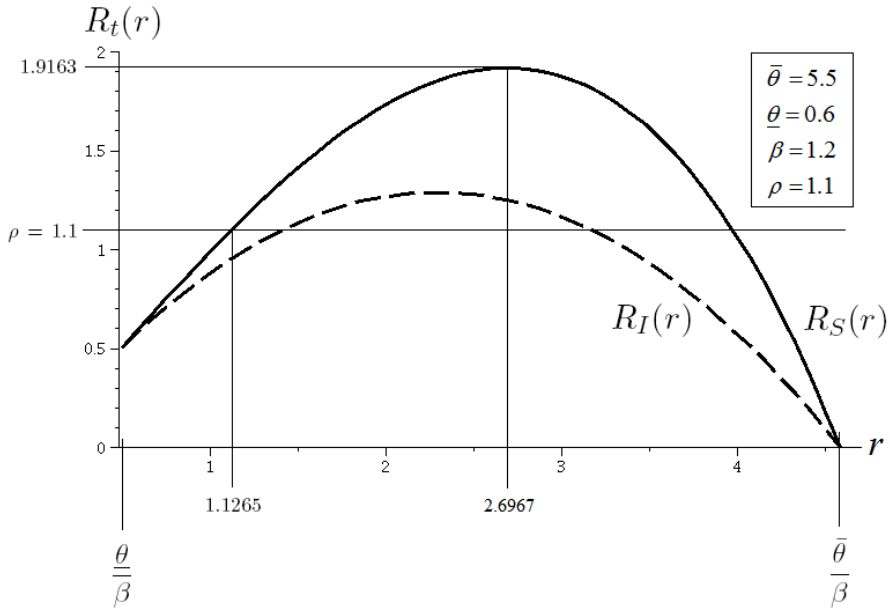


Figure 3.16: Expected repayments with social sanctions.

4. We have shown that the equilibrium loan rate is $r_S \leq r_S^{max}$. It follows that $q = m$, i.e., all borrowers demand loans, since $U_S(r) > 0$ for (not only, but in particular) all $\frac{\underline{\theta}}{\beta} < r \leq r_S^{max}$ (cf. Figure 3.15). q.e.d.

Example 1 (ctd.): In our example with $\underline{\theta} = 0.6$, $\bar{\theta} = 5.5$, $\beta = 1.2$, and $\rho = 1.1$, the equilibrium loan rate is $r_S = 1.1265$. The repayment rate and expected utility rise to 97.65% and 1.9203, respectively. $R_S(r)$ achieves its maximum $R_S^{max} = 1.9163$ at $r_S^{max} = 2.6967$, so that an equilibrium exists for costs of capital up to 1.9163 here (see Figure 3.16).

3.6.3 Discussion of social sanctions

Let us come back to the fact that $U_S(r)$ might be negative, as depicted in the lower panel of Figure 3.15. We have already said that it is the use of second-period income to repay which might make projects under GL with social sanctions unattractive. In the left panel of Figure 3.17, we have a rather large loan rate and some positive β. In the shaded area (AB-1), project payoffs are below the loan rate for one borrower, but social sanctions make both

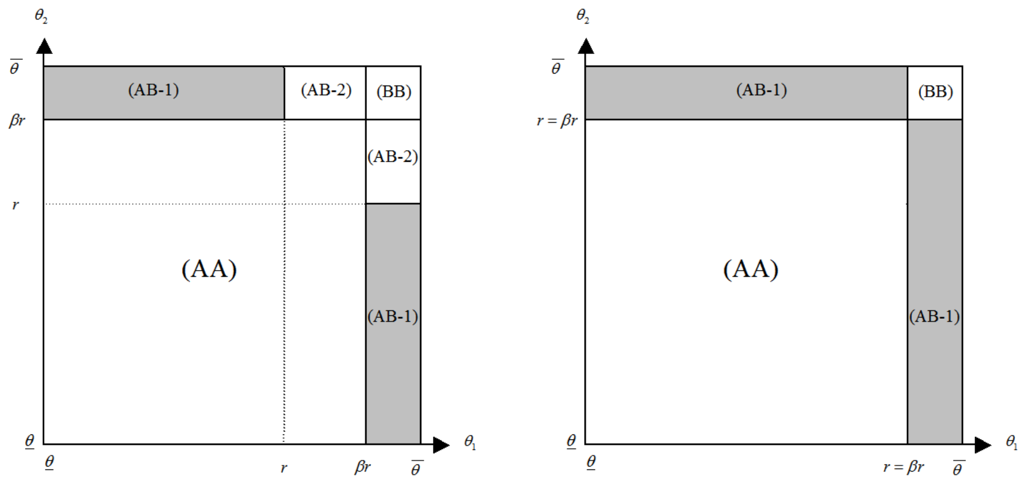


Figure 3.17: Cases of negative expected utility, left panel: $\beta > 1$, right panel: $\beta = 1$.

borrowers contribute r, one of the two using alternative, second-period income (borrower 1 in the shaded area in the upper left). We construct a situation with negative expected utility. Let β approach one (see the right panel in Figure 3.17). Thus, in the limit, $\beta r \Rightarrow r$, and area (AB-2) disappears. Since there is default in area (AA), borrowers' actual net payoff is $\theta - p(\theta) = \theta - \theta = 0$ in that area. In area (BB), both contribute their share and are left with some small positive net payoff. In area (AB-1), one borrower has a tiny positive payoff, whereas the other might have to use an immense amount of second-period income to contribute r. Thus, parameter constellations with β sufficiently close to one imply the existence of rather high loan rates such that borrowers' expected utility is negative.[53] However, a result of the proof of Proposition 3.7 is that these loan rates are not offered by banks.

Proposition 3.7 shows that the disadvantage of GL, which potentially makes IL the equilibrium mode of finance despite the higher break-even loan rate (cf. Proposition 3.2), can be overcome by means of social sanctions: With social sanctions obeying Assumption 6, lending takes place using a GL contract whenever projects are such that banks can break even at some loan rate. This does not mean, however, that GL helps to get rid of the market failures due to enforcement problems altogether: The fact that expected repayment $R_S(r)$ is hump-shaped implies that the allocation problems encountered in Subsections 3.4.4 - 3.4.6 continue

[53]For instance, using $\beta = 1.01$, $\underline{\theta} = 0.6$, $\bar{\theta} = 5.5$, and $\alpha = 0$ yields a negative expected utility for all $3.67 < r < 5.38$.

to be prevalent. That is, there is financial fragility in that the market collapses when ρ rises beyond R_S^{max}; if there are several borrower classes j, those with $R_S^{max,j} < \rho$ are redlined; and if capital supply is a strictly increasing function $s(\rho)$, credit rationing arises if $s(R_S^{max}) < m$ (cf. Propositions 3.3-3.5).

Moreover, the use of social sanctions is not without cost: Borrowers might end up in deeper poverty if they accept a group loan. The existence of social sanctions might force them to sell their belongings in order to avoid peer punishment. Our proof that expected utility is always positive in any potential equilibrium with GL and social sanctions crucially hinges on two assumptions.

First, and more importantly, we assumed that borrowers are risk-neutral. However, our motivation to introduce exogenous second-period income in the first place was to account for the fact that borrowers could even sell the last of the little they have. Thus, it would be more appropriate to assume loss aversion regarding second-period income. The perceived loss for a borrower who becomes homeless probably outweighs his perceived increase in utility from being able to afford new trousers. It is not difficult to see how our model yields negative expected utility if borrowers are loss-averse. In its most extreme form,[54] we could assume an infinite negative payoff from having to use exogenous second-period income.[55] In that case, there cannot be an equilibrium with GL contracts at loan rates such that the probability for case (AB-1) is positive, since demand is zero.[56]

Second, we have assumed that penalties are non-pecuniary. This is in line with most of the literature on (optimal) GL contracts. Our results confirm what most of the studies have found: When social sanctions are in place, GL improves on IL, either in terms of repayment rates or borrower utility. However, let us repeat that BC point out that penalties are partly "a monetary loss due to seizure of income or assets" (p.4). In our model, using $\alpha > 0$ can lead to negative expected utility of borrowers at loan rates that banks would offer. This can most easily be seen setting $\alpha = 1$ and doing some numerical calculations. For instance, using $\rho = 1.4$, $\underline{\theta} = 0.6$, $\bar{\theta} = 2.1$ $\beta = 1.01$ and $\alpha = 1$, expected utility becomes negative at r_S, the former equilibrium loan rate.

[54] Recall the bank's infinite negative payoff from default, which we used in Chapter 2.
[55] Note that sanctions must then also be infinitely high to leave repayment behavior in a group unchanged.
[56] As a consequence, given that $R_S^{max} > \rho$, the credit market might break down if $R_I^{max} < \rho$.

3.7 Conclusion

BC analyze repayment rates in a GL model with enforcement problems. The recent trend toward private investments and market instruments in microfinance markets raises the question of what equilibrium in the BC model looks like. This chapter shows that the joint liability feature of GL in combination with non-cooperative behavior of borrowers does not generally make GL the equilibrium mode of finance (which yields the higher borrower utility). We have shown how cooperation eliminates the surprising result that GL might have the higher repayment rate, be feasible at the lower contractual loan rate, but nevertheless not be the equilibrium mode of finance. More than that, we have shown that GL is the unique equilibrium mode of finance when borrowers cooperate in the repayment game and penalties are non-pecuniary. When borrowers use social sanctions, non-pecuniary penalties also make GL the unique equilibrium mode of finance if borrowers are risk-neutral both regarding project payoffs and with regard to what they possess. In particular, even though borrowers might be worse-off from taking a loan ex post, they always demand loans ex ante. The reason is that all loan rates MFIs can offer (given competition) are such that borrowers' expected utility from borrowing is positive. This is not necessarily true if penalties are partly pecuniary or if borrowers are risk-averse regarding the loss of their belongings.

Irrespective of the type of contract used, and, if it is GL, irrespective of the existence of social sanctions and of whether borrowers cooperate or not, the market equilibrium suffers from the usual allocation problems known from the imperfect information literature. This means that the prospective growth of the market for microcredit is unlikely to be a frictionless process. MFIs will have to continue to take due care that borrowers have proper incentives to repay. If the DFIs' objective is to maximize the loan volume, they should target MFIs active in the less profitable segments of the market and leave the more profitable business to private investors.

Even though economic analysis is an important source of scientific knowledge, we want to mention the importance of results from adjacent sciences, in particular when they concern phenomena in communities where incentives are strongly based on social capital. For instance, the moral hazard literature assumes that borrowers can influence project payoffs, which is usually modeled as an increase in the projects' success probability. The terminology of the economic literature seems to suggest that borrowers are intrinsically bad and shirk if not properly monitored or threatened with sanctions. Thus, the performance of group lending is frequently attributed to 'peer monitoring' or 'peer sanctioning'. However, based on our reading on group lending in general and the performance of groups in particular, we want to stress some results from psychology and sociology. Being part of a 'greater something' (like a lending group) might spur productivity (and well-being). It is not only the fear of sanctions that makes borrowers work hard, but the desire to contribute to the success of the group. 'Identity theory' and its focus on individualistic outcomes of identity-related processes (to be distinguished from 'social identity theory') can help explain the idea. In an attempt to describe differences and commonalities of both theories, Hogg, Terry, and White (1995, p.257) contend that "[t]he perception that one is enacting a role satisfactorily should enhance feelings of self-esteem, whereas perceptions of poor role performance may engender doubts about one's self-worth, and may even produce symptoms of psychological distress". We are convinced that the link to increased productivity is not too far-fetched. In their analysis 'social identity and individual productivity within groups', Worchel, Rothgerber, Day, Hart, and Butemeyer (1998) confirm this link. Many studies find a positive influence of group lending on project revenues. For instance, Gomez and Santor (2003, p.v) find that borrowers in a group spend more effort than their individual lending counterparts. Clearly, it is hard, if not impossible, to disentangle the effects on effort and productivity that come from extrinsic incentives like the existence of sanctions from those that arise due to our postulated intrinsic channel based on 'identity theory'.[57] The distinction between intrinsic motivation and extrinsic incentives is not new, but increasingly recognized as an important area of economic research. Kreps (1997, p.363) contends that this will "involve activities unfamiliar to economics", which are nonetheless "important and must be pursued".

In this chapter, we have focused on the last mile of microcredit. However, we partly motivated our investigation from the capital supply side. Reille, Glisovic-Mezieres, Berthouzoz, and Milverton (2009, p.1) claim that "[f]oreign capital investments in microfinance passed the US $10 billion mark in December 2008. More than half of this cross-border investment is

[57] In a first stage, one would have to make sure that differences in productivity do not stem from selection effects: More productive borrowers might choose group lending, see the study of Van Tassel (1999), who analyzes this point for different risk types.

managed by MIVs. This new and fast-growing segment of the emerging market asset class is attracting a broad range of socially oriented investors." This naturally poses the question of how socially oriented investors behave. If they do not maximize financial returns, there must be some other approach to take investment and portfolio decisions, probably based on some social dimension. Can the behavior of socially oriented investors be quantified? Morduch (1999, p.1572) writes that "[t]he promise of MF was founded on innovation: new management structures, new contracts, and new attitudes". This chapter has focused on new management techniques and contracts. The next chapter will consider new attitudes in the particular context of portfolio choice.

3.8 Appendix

3.8.1 Expected repayment and expected utility in cases L and H

1. Expected repayment and expected utility in case L:

The expected penalty conditional on default is necessary to calculate expected repayment (cf. equation (3.4)). Since $p(\theta) = \frac{\theta}{\beta}$, the expected penalty is expected payoff divided by β. Expected payoff conditional on default is the sum of expected payoffs conditional on $\theta \in A$ and $\theta \in B$, respectively, weighted by their respective probabilities

$$\begin{aligned}
E[\theta|\,\boldsymbol{\theta}_G \in D_G] &= \frac{1}{1-\Pi_G(r)}\left(\int_{\underline{\theta}}^{2\beta r}\frac{1}{\bar{\theta}-\underline{\theta}}\int_{\underline{\theta}}^{\beta r}\frac{\theta}{\bar{\theta}-\underline{\theta}}d\theta\, d\theta' + \int_{\underline{\theta}}^{\beta r}\frac{1}{\bar{\theta}-\underline{\theta}}\int_{\beta r}^{2\beta r}\frac{\theta}{\bar{\theta}-\underline{\theta}}d\theta\, d\theta'\right)\\
&= \frac{1}{1-\Pi_G(r)}\left[\frac{F(2\beta r)}{\bar{\theta}-\underline{\theta}}\int_{\underline{\theta}}^{\beta r}\theta\, d\theta + \frac{F(\beta r)}{\bar{\theta}-\underline{\theta}}\int_{\beta r}^{2\beta r}\theta\, d\theta\right]\\
&= \frac{1}{1-\Pi_G(r)}\left[\frac{2\beta r-\underline{\theta}}{(\bar{\theta}-\underline{\theta})^2}\frac{\beta^2 r^2-\underline{\theta}^2}{2} + \frac{\beta r-\underline{\theta}}{(\bar{\theta}-\underline{\theta})^2}\frac{3\beta^2 r^2}{2}\right]\\
&= \frac{1}{1-\Pi_G(r)}\frac{5\beta^3 r^3 - 4\beta^2\underline{\theta} r^2 - 2\beta\underline{\theta}^2 r + \underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2}.
\end{aligned}$$

Using this together with $\Pi_G(r)$ from equation (3.2), expected repayment becomes

$$\begin{aligned}
R_G(r) &= \Pi_G(r)r + [1-\Pi_G(r)]\alpha E\left(\frac{\theta}{\beta}\middle|\boldsymbol{\theta}_G \in D_G\right)\\
&= \frac{1}{2(\bar{\theta}-\underline{\theta})^2}\left[(-6\beta^2 r^2 + 8\beta\underline{\theta} r + 2\bar{\theta}^2 - 4\underline{\theta}\bar{\theta})r + \frac{\alpha}{\beta}(5\beta^3 r^3 - 4\beta^2\underline{\theta} r^2 - 2\beta\underline{\theta}^2 r + \underline{\theta}^3)\right]\\
&= \frac{-\beta^2(6-5\alpha)r^3 + 4\beta\underline{\theta}(2-\alpha)r^2 + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta} - \alpha\underline{\theta}^2)r + \frac{\alpha\underline{\theta}^3}{\beta}}{2(\bar{\theta}-\underline{\theta})^2}.
\end{aligned}$$

To derive expected utility as in equation (3.6), we need to know both $E[\theta|\,\boldsymbol{\theta}_G \in D_G]$ and $\Pi_G(r)E[\theta - r|\,\boldsymbol{\theta}_G \in S_G]$. We have already found the former. Since

$$\frac{\bar{\theta}+\underline{\theta}}{2} = E[\theta] = \Pi_G(r)E[\theta|\,\boldsymbol{\theta}_G \in S_G] + (1-\Pi_G(r))E[\theta|\,\boldsymbol{\theta}_G \in D_G], \tag{3.30}$$

the latter is

$$\begin{aligned}
\Pi_G(r)E[\theta|\,\boldsymbol{\theta}_G \in S_G] &= \frac{\bar{\theta}+\underline{\theta}}{2} - (1-\Pi_G(r))E[\theta|\,\boldsymbol{\theta}_G \in D_G] \\
&= \frac{\bar{\theta}+\underline{\theta}}{2} - \frac{5\beta^3r^3 - 4\beta^2\underline{\theta}r^2 - 2\beta\underline{\theta}^2r + \underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{-5\beta^3r^3 + 4\beta^2\underline{\theta}r^2 + 2\beta\underline{\theta}^2r + \bar{\theta}^3 - \underline{\theta}\bar{\theta}^2 - \underline{\theta}^2\bar{\theta}}{2(\bar{\theta}-\underline{\theta})^2}.
\end{aligned}$$

Expected utility becomes

$$\begin{aligned}
U_G(r) &= \Pi_G(r)E[\theta - r|\,\boldsymbol{\theta}_G \in S_G] + (1-\Pi_G(r))E\left[\theta - \frac{\theta}{\beta}\middle|\,\boldsymbol{\theta}_G \in D_G\right] \\
&= \frac{-5\beta^3r^3 + 4\beta^2\underline{\theta}r^2 + 2\beta\underline{\theta}^2r + \bar{\theta}^3 - \underline{\theta}\bar{\theta}^2 - \underline{\theta}^2\bar{\theta}}{2(\bar{\theta}-\underline{\theta})^2} \\
&\quad - \frac{-3\beta^2r^2 + 4\beta\underline{\theta}r + \bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{(\bar{\theta}-\underline{\theta})^2} r \\
&\quad + \left(1 - \frac{1}{\beta}\right)\frac{5\beta^3r^3 - 4\beta^2\underline{\theta}r^2 - 2\beta\underline{\theta}^2r + \underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{\beta^2r^3 - 4\beta\underline{\theta}r^2 + (2\underline{\theta}^2 - 2\bar{\theta}^2 + 4\underline{\theta}\bar{\theta})r + \bar{\theta}^3 - \underline{\theta}\bar{\theta}^2 - \underline{\theta}^2\bar{\theta} + \left(1-\frac{1}{\beta}\right)\underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2}.
\end{aligned}$$

2. Expected repayment and expected utility in case H:

In analogy to case L, we calculate expected payoffs conditional on default.

$$\begin{aligned}
E[\theta|\,\boldsymbol{\theta}_G \in D_G] &= \frac{1}{1-\Pi_G(r)}\left(\int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{\bar{\theta}-\underline{\theta}} \int_{\underline{\theta}}^{\beta r} \frac{\theta}{\bar{\theta}-\underline{\theta}} d\theta\, d\theta' + \int_{\underline{\theta}}^{\beta r} \frac{1}{\bar{\theta}-\underline{\theta}} \int_{\beta r}^{\bar{\theta}} \frac{\theta}{\bar{\theta}-\underline{\theta}} d\theta\, d\theta'\right) \\
&= \frac{1}{1-\Pi_G(r)}\left[\frac{1}{\bar{\theta}-\underline{\theta}}\int_{\underline{\theta}}^{\beta r} \theta\, d\theta + \frac{F(\beta r)}{\bar{\theta}-\underline{\theta}}\int_{\beta r}^{\bar{\theta}} \theta\, d\theta\right] \\
&= \frac{1}{1-\Pi_G(r)}\left[\frac{1}{\bar{\theta}-\underline{\theta}}\frac{\beta^2r^2 - \underline{\theta}^2}{2} + \frac{\beta r - \underline{\theta}}{(\bar{\theta}-\underline{\theta})^2}\frac{\bar{\theta}^2 - \beta^2r^2}{2}\right] \\
&= \frac{1}{1-\Pi_G(r)}\frac{-\beta^3r^3 + \beta^2\bar{\theta}r^2 + \beta\bar{\theta}^2r - \underline{\theta}^2\bar{\theta} - \underline{\theta}\bar{\theta}^2 + \underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2}.
\end{aligned}$$

Using the above and $\Pi_G(r)$ from equation (3.3), we get expected repayment:

$$
\begin{aligned}
R_G(r) &= \Pi_G(r)r + [1-\Pi_G(r)]\alpha E\left(\frac{\theta}{\beta}\middle|\boldsymbol{\theta}_G \in D_G\right) \\
&= \left(\frac{\bar{\theta}-\beta r}{\bar{\theta}-\underline{\theta}}\right)^2 r + \frac{\alpha}{\beta}\frac{-\beta^3 r^3 + \beta^2\bar{\theta}r^2 + \beta\bar{\theta}^2 r - \underline{\theta}^2\bar{\theta} - \underline{\theta}\bar{\theta}^2 + \underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{\beta^2(2-\alpha)r^3 - \beta\bar{\theta}(4-\alpha)r^2 + \bar{\theta}^2(2+\alpha)r - \frac{\alpha}{\beta}(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2}.
\end{aligned}
$$

Equation (3.30) is valid in case H, too. Using $\Pi_G(r)$ from equation (3.3), we have

$$
\begin{aligned}
\Pi_G(r)E[\theta|\,\boldsymbol{\theta}_G \in S_G] &= \frac{\bar{\theta}+\underline{\theta}}{2} - (1-\Pi_G(r))E[\theta|\,\boldsymbol{\theta}_G \in D_G] \\
&= \frac{\bar{\theta}+\underline{\theta}}{2} - \frac{-\beta^3 r^3 + \beta^2\bar{\theta}r^2 + \beta\bar{\theta}^2 r - \underline{\theta}^2\bar{\theta} - \underline{\theta}\bar{\theta}^2 + \underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{\beta^3 r^3 - \beta^2\bar{\theta} - \beta\bar{\theta}^2 r + \bar{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2},
\end{aligned}
$$

so that expected utility becomes

$$
\begin{aligned}
U_G(r) &= \Pi_G(r)E[\theta - r|\,\boldsymbol{\theta}_G \in S_G] + [1-\Pi_G(r)]E\left[\theta - \frac{\theta}{\beta}\middle|\boldsymbol{\theta}_G \in D_G\right] \\
&= \frac{\beta^3 r^3 - \beta^2\bar{\theta}r^2 - \beta\bar{\theta}^2 r + \bar{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} - \left(\frac{\bar{\theta}-\beta r}{\bar{\theta}-\underline{\theta}}\right)^2 r \\
&\quad + \left(1-\frac{1}{\beta}\right)\frac{-\beta^3 r^3 + \beta^2\bar{\theta}r^2 + \beta\bar{\theta}^2 r - \underline{\theta}^2\bar{\theta} - \underline{\theta}\bar{\theta} + \underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{\bar{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} - \frac{2\bar{\theta}^2 r - 4\beta\bar{\theta}r^2 + 2\beta^2 r^3}{2(\bar{\theta}-\underline{\theta})^2} \\
&\quad + \frac{-\underline{\theta}^2\bar{\theta} - \underline{\theta}\bar{\theta}^2 + \underline{\theta}^3 + \beta^2 r^3 - \beta\bar{\theta}r^2 - \bar{\theta}^2 r + \frac{1}{\beta}(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{-\beta^2 r^3 + 3\beta\bar{\theta}r^2 - 3\bar{\theta}^2 r + \bar{\theta}^3 - \left(1-\frac{1}{\beta}\right)(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2}.
\end{aligned}
$$

3.8.2 Proof: $R_G(r)$ continuous at $\frac{\bar{\theta}}{2\beta}$

Expected repayment at $r = \frac{\bar{\theta}}{2\beta}$, the threshold between cases L and H, is

$$R_G\left(\frac{\bar{\theta}}{2\beta}\right)_L = \frac{-\beta^2(6-5\alpha)(\frac{\bar{\theta}}{2\beta})^3 + 4\beta\underline{\theta}(2-\alpha)(\frac{\bar{\theta}}{2\beta})^2 + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta} - \alpha\underline{\theta}^2)\frac{\bar{\theta}}{2\beta} + \frac{\alpha\underline{\theta}^3}{\beta}}{2(\bar{\theta}-\underline{\theta})^2},$$

for case L, and

$$R_G\left(\frac{\bar{\theta}}{2\beta}\right)_H = \frac{\beta^2(2-\alpha)(\frac{\bar{\theta}}{2\beta})^3 - \beta\bar{\theta}(4-\alpha)(\frac{\bar{\theta}}{2\beta})^2 + \bar{\theta}^2(2+\alpha)\frac{\bar{\theta}}{2\beta} - \frac{\alpha}{\beta}(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2},$$

for case H. We have to show that the difference between both expressions is zero. Multiplying by $2\beta(\bar{\theta}-\underline{\theta})^2$ and rearranging yields

$$\begin{aligned}
R_G\left(\frac{\bar{\theta}}{2\beta}\right)_L - R_G\left(\frac{\bar{\theta}}{2\beta}\right)_H &= -\frac{6\bar{\theta}^3}{8} + \frac{5\alpha\bar{\theta}^3}{8} + 2\bar{\theta}^2\underline{\theta} - \alpha\bar{\theta}^2\underline{\theta} + \bar{\theta}^3 - 2\bar{\theta}^2\underline{\theta} - \alpha\underline{\theta}^2\bar{\theta} + \alpha\underline{\theta}^3 - \frac{\bar{\theta}^3}{4} + \frac{\alpha\bar{\theta}^3}{8} \\
&\quad + \bar{\theta}^3 - \frac{\alpha\bar{\theta}^3}{4} - \bar{\theta}^3 - \frac{\alpha\bar{\theta}^3}{2} + \alpha\bar{\theta}\underline{\theta}^2 + \alpha\bar{\theta}^2\underline{\theta} - \alpha\underline{\theta}^3 \\
&= \bar{\theta}^3\left(-\frac{3}{4} - \frac{1}{4} + \frac{5\alpha}{8} + \frac{\alpha}{8} + 1 - \frac{\alpha}{4} - \frac{\alpha}{2}\right) \\
&= 0.
\end{aligned}$$

3.8.3 Intersections of $R_I(r)$ and $R_G(r)$

1. Proof that $\boldsymbol{R_I(\underline{\theta}/\beta) = R_G(\underline{\theta}/\beta) = \underline{\theta}/\beta}$:

For IL, inserting $r = \frac{\underline{\theta}}{\beta}$ into equation (3.1) yields

$$R_I\left(\frac{\underline{\theta}}{\beta}\right) = \frac{-\frac{\underline{\theta}^2}{\beta} + 2\frac{\underline{\theta}\bar{\theta}}{\beta} - \frac{\underline{\theta}^2}{\beta}}{2(\bar{\theta}-\underline{\theta})} = \frac{\frac{\underline{\theta}}{\beta}(\bar{\theta}-\underline{\theta})}{\bar{\theta}-\underline{\theta}} = \frac{\underline{\theta}}{\beta}.$$

For GL, inserting $r = \frac{\underline{\theta}}{\beta}$ (case L) into equation (3.10) yields

$$\begin{aligned}
R_G\left(\frac{\underline{\theta}}{\beta}\right) &= \frac{-\beta^2(6-5\alpha)\frac{\underline{\theta}^3}{\beta^3}+4\beta\underline{\theta}(2-\alpha)\frac{\underline{\theta}^2}{\beta^2}+2(\bar{\theta}^2-2\underline{\theta}\bar{\theta}-\alpha\underline{\theta}^2)\frac{\underline{\theta}}{\beta}+\alpha\frac{\underline{\theta}^3}{\beta}}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{2\frac{\underline{\theta}^3}{\beta}+2\frac{\underline{\theta}}{\beta}(\bar{\theta}^2-2\underline{\theta}\bar{\theta})}{2(\bar{\theta}-\underline{\theta})^2} = \frac{\underline{\theta}}{\beta}\frac{\underline{\theta}^2-2\underline{\theta}\bar{\theta}+\bar{\theta}^2}{(\bar{\theta}-\underline{\theta})^2} = \frac{\underline{\theta}}{\beta}.
\end{aligned}$$

2. Proof that $R_I(\bar{\theta}/\beta) = R_G(\bar{\theta}/\beta) = (\alpha/\beta)(\bar{\theta}+\underline{\theta})/2$:

For IL, inserting $r = \frac{\bar{\theta}}{\beta}$ into equation (3.1) yields

$$R_I\left(\frac{\bar{\theta}}{\beta}\right) = \frac{-\beta(2-\alpha)\left(\frac{\bar{\theta}}{\beta}\right)^2+2\bar{\theta}\frac{\bar{\theta}}{\beta}-\alpha\frac{\underline{\theta}^2}{\beta}}{2(\bar{\theta}-\underline{\theta})} = \frac{\alpha\frac{\bar{\theta}^2}{\beta}-\alpha\frac{\underline{\theta}^2}{\beta}}{2(\bar{\theta}-\underline{\theta})} = \frac{\alpha}{\beta}\frac{\bar{\theta}+\underline{\theta}}{2}.$$

For GL, inserting $r = \frac{\bar{\theta}}{\beta}$ (case H) into equation (3.12) yields

$$\begin{aligned}
R_G\left(\frac{\bar{\theta}}{\beta}\right) &= \frac{\beta^2(2-\alpha)\left(\frac{\bar{\theta}}{\beta}\right)^3+\beta\bar{\theta}(4-\alpha)\left(\frac{\bar{\theta}}{\beta}\right)^2+\bar{\theta}^2(2+\alpha)\frac{\bar{\theta}}{\beta}-\frac{\alpha}{\beta}(\underline{\theta}^2\bar{\theta}+\underline{\theta}\bar{\theta}^2-\underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{(2-\alpha)\frac{\bar{\theta}^3}{\beta}-(4-\alpha)\frac{\bar{\theta}^3}{\beta}+(2+\alpha)\frac{\bar{\theta}^3}{\beta}-\frac{\alpha}{\beta}(\underline{\theta}^2\bar{\theta}+\underline{\theta}\bar{\theta}^2-\underline{\theta}^3)}{2(\bar{\theta}-\underline{\theta})^2} \\
&= \frac{\alpha}{\beta}\frac{\bar{\theta}^3-\underline{\theta}^2\bar{\theta}-\underline{\theta}\bar{\theta}^2+\underline{\theta}^3}{2(\bar{\theta}-\underline{\theta})^2} = \frac{\alpha}{\beta}\frac{\bar{\theta}^2(\bar{\theta}-\underline{\theta})-\underline{\theta}^2(\bar{\theta}-\underline{\theta})}{2(\bar{\theta}-\underline{\theta})^2} = \frac{\alpha}{\beta}\frac{\bar{\theta}^2-\underline{\theta}^2}{2(\bar{\theta}-\underline{\theta})} = \frac{\alpha}{\beta}\frac{\bar{\theta}+\underline{\theta}}{2}.
\end{aligned}$$

3.8.4 Proof that $U_t'(r) < 0$ for $r < \bar{\theta}/\beta$, $t \in \{I, G\}$

1. Individual lending

From (3.9),

$$U_I'(r) = \frac{\beta r-\bar{\theta}}{\bar{\theta}-\underline{\theta}} < 0 \Leftrightarrow r < \frac{\bar{\theta}}{\beta}.$$

2. Group lending

a) Case L

Let $U_{GL}(r)$ denote the function on the right-hand side of (3.11). Differentiating with respect to r yields

$$\begin{aligned} U'_{GL}(r) &= \frac{3\beta^2 r^2 - 8\beta\underline{\theta} r + (2\underline{\theta}^2 - 2\bar{\theta}^2 + 4\underline{\theta}\bar{\theta})}{2(\bar{\theta}-\underline{\theta})^2} = \frac{3\beta^2\left(r^2 - \frac{8}{3}\frac{\underline{\theta}}{\beta}r + \frac{2}{3}\frac{\underline{\theta}^2 - \bar{\theta}^2 + 2\underline{\theta}\bar{\theta}}{\beta^2}\right)}{2(\bar{\theta}-\underline{\theta})^2} \\ &< 0 \Leftrightarrow 0 > r^2 - \frac{8}{3}\frac{\underline{\theta}}{\beta}r + \frac{2}{3}\frac{\underline{\theta}^2 - \bar{\theta}^2 + 2\underline{\theta}\bar{\theta}}{\beta^2}. \end{aligned}$$

Let $x \equiv \bar{\theta}/(2\underline{\theta}) > 1$. The roots of the polynomial on the right-hand side of the last inequality are

$$\begin{aligned} r_{1/2} &= \frac{4}{3}\frac{\underline{\theta}}{\beta} \pm \left\{ \frac{16}{9}\left(\frac{\underline{\theta}}{\beta}\right)^2 - \frac{2}{3}\left(\frac{\underline{\theta}}{\beta}\right)^2 \left[1 - \left(\frac{\bar{\theta}}{\underline{\theta}}\right)^2 + 2\frac{\bar{\theta}}{\underline{\theta}}\right] \right\}^{\frac{1}{2}} \\ &= \frac{4}{3}\frac{\underline{\theta}}{\beta}\left[1 \pm \left(\frac{3}{2}x^2 - \frac{3}{2}x + \frac{5}{8}\right)^{\frac{1}{2}}\right]. \end{aligned}$$

Since $x > 1$, the discriminant is positive, so there are two real roots. The smaller root r_1 is less than $\underline{\theta}/\beta$ iff

$$\begin{aligned} \frac{4}{3}\frac{\underline{\theta}}{\beta}\left[1 - \left(\frac{3}{2}x^2 - \frac{3}{2}x + \frac{5}{8}\right)^{\frac{1}{2}}\right] &< \frac{\underline{\theta}}{\beta} \\ \Leftrightarrow \frac{4}{3}\left[1 - \left(\frac{3}{2}x^2 - \frac{3}{2}x + \frac{5}{8}\right)^{\frac{1}{2}}\right] &< 1. \end{aligned}$$

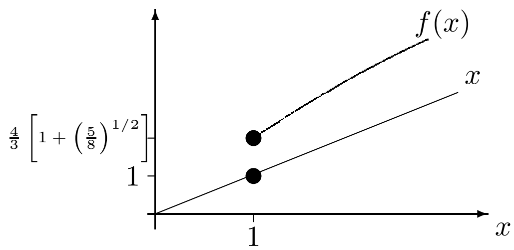
Since $x > 1$, a sufficient condition for the validity of this inequality is

$$1 > \frac{4}{3}\left[1 - \left(\frac{5}{8}\right)^{\frac{1}{2}}\right] \approx 0.2792.$$

The bigger root r_2 is greater than $\bar{\theta}/(2\beta)$ iff

$$\frac{4}{3}\frac{\underline{\theta}}{\beta}\left[1+\left(\frac{3}{2}x^2-\frac{3}{2}x+\frac{5}{8}\right)^{\frac{1}{2}}\right] > \frac{\bar{\theta}}{2\beta} \Leftrightarrow x < \frac{4}{3}\left[1+\left(\frac{3}{2}x^2-\frac{3}{2}x+\frac{5}{8}\right)^{\frac{1}{2}}\right] \equiv f(x). \tag{3.31}$$

The function $f(x)$ is larger than x for all $x \geq 1$:



This follows from the fact that

$$f(1) = \frac{4}{3}\left[1+\left(\frac{5}{8}\right)^{\frac{1}{2}}\right] > 1,$$

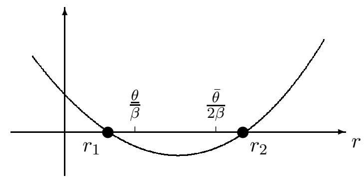
and $f'(x) > 1$ for all $x \geq 1$:

$$\begin{aligned}
f'(x) &= \frac{4}{3}\left(3x-\frac{3}{2}\right)\frac{1}{2}\left(\frac{3}{2}x^2-\frac{3}{2}x+\frac{5}{8}\right)^{-\frac{1}{2}} \\
&> 1 \Leftrightarrow \\
12x-6 &> 6\left(\frac{3}{2}x^2-\frac{3}{2}x+\frac{5}{8}\right)^{\frac{1}{2}} \\
2x-1 &> \left(\frac{3}{2}x^2-\frac{3}{2}x+\frac{5}{8}\right)^{\frac{1}{2}} \\
4x^2-4x+1 &> \frac{3}{2}x^2-\frac{3}{2}x+\frac{5}{8} \\
\frac{5}{2}x(x-1)+\frac{3}{8} &> 0.
\end{aligned}$$

So the r-values consistent with the definition of case L form a subset of the r-values such that $U'_{GL}(r) < 0$:

$$\left[\frac{\underline{\theta}}{\beta}, \frac{\bar{\theta}}{2\beta}\right] \subset (r_1, r_2).$$

The following graph shows $U'_{GL}(r)$:



This proves $U'_G(r) < 0$ in case L.

b) case H

Letting $U_{GH}(r)$ denote the function on the right-hand side of (3.13), we have:

$$\begin{aligned} U'_{GH}(r) &= \frac{-3\beta^2 r^2 + 6\beta\bar{\theta}r - 3\bar{\theta}^2}{2(\bar{\theta} - \underline{\theta})^2} \\ &= \frac{-3\beta^2 \left[r^2 - 2\frac{\bar{\theta}}{\beta} r + \left(\frac{\bar{\theta}}{\beta}\right)^2 \right]}{2(\bar{\theta} - \underline{\theta})^2} \\ &= \frac{-3\beta^2 \left(r - \frac{\bar{\theta}}{\beta} \right)^2}{2(\bar{\theta} - \underline{\theta})^2} < 0. \end{aligned}$$

So $U'_G(r) < 0$ in case H, too.

3.8.5 Proof: maximum expected repayment in example 1.

Using equation (3.34) from Appendix 3.8.8 below, we get

$$R'_{GL0}\left(\frac{\bar{\theta}}{2\beta}\right) = -\frac{\frac{5}{2}\bar{\theta}\left(\bar{\theta} - \frac{8}{5}\underline{\theta}\right)}{2(\bar{\theta} - \underline{\theta})^2} < 0$$

(recall that we assume $\bar{\theta}/2 > \underline{\theta}$). From Appendix 3.8.8 below, we know that $R'_{GL0}(\bar{\theta}/(3\beta)) > 0$ so that $R_{GL0}(r)$ attains a maximum between $r = \bar{\theta}/(3\beta)$ and $r = \bar{\theta}/(2\beta)$. The maximizing r-value can be calculated explicitly:

$$
\begin{aligned}
0 &= R'_{GL0}(r) \\
0 &= -18\beta^2 r^2 + 16\beta\underline{\theta} r + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta}) \\
0 &= r^2 - \frac{8}{9}\frac{\underline{\theta}}{\beta} r - \frac{\bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{9\beta^2} \\
r_{1/2} &= \frac{4}{9}\frac{\underline{\theta}}{\beta} \pm \left[\frac{1}{4}\left(\frac{8}{9}\frac{\underline{\theta}}{\beta}\right)^2 + \frac{\bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{9\beta^2}\right]^{\frac{1}{2}} = \frac{4}{9}\frac{\underline{\theta}}{\beta}\left[1 \pm \left(1 + \frac{9}{16}\frac{\bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{\underline{\theta}^2}\right)^{\frac{1}{2}}\right].
\end{aligned}
$$

The smaller root is negative, the larger positive: $r_1 < 0 < r_2$. Using parameters as in example 1, viz., $\bar{\theta} = 5.5$, $\underline{\theta} = 0.6$, and $\beta = 1.2$, $r_2 \approx 1.5912$. Inserting this into $R_{GL0}(r)$ in equation (3.32) gives a maximum expected repayment under GL of 1.1462.

3.8.6 Proof: $R_G(r) > R_I(r)$ in L,H when $\alpha = 1$

We want to show that expected repayment under GL is higher than expected repayment under IL if penalties are 100 % pecuniary. The range of loan rates we have to consider is $\underline{\theta}/\beta < r < \bar{\theta}/\beta$. We prove this property separately for cases L and H.

a) Case L

Let $R_{I1}(r)$ and $R_{GL1}(r)$ denote the functions on the right-hand sides of (3.8) and (3.10) for $\alpha = 1$, respectively:

$$
\begin{aligned}
R_{I1}(r) &= \frac{-\beta r^2 + 2\bar{\theta} r - \frac{\underline{\theta}^2}{\beta}}{2(\bar{\theta} - \underline{\theta})}, \\
R_{GL1}(r) &= \frac{-\beta^2 r^3 + 4\beta\underline{\theta} r^2 + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta} - \underline{\theta}^2) r + \frac{\underline{\theta}^3}{\beta}}{2(\bar{\theta} - \underline{\theta})^2}.
\end{aligned}
$$

As shown in Appendix 3.8.3,

$$
R_{I1}\left(\frac{\underline{\theta}}{\beta}\right) = R_{GL1}\left(\frac{\underline{\theta}}{\beta}\right) = \frac{\underline{\theta}}{\beta}.
$$

Differentiating $R_{I1}(r)$ and $R_{GL1}(r)$ with respect to r gives

$$\begin{aligned} R'_{I1}(r) &= \frac{-2\beta r + 2\bar{\theta}}{2(\bar{\theta} - \underline{\theta})}, \\ R'_{GL1}(r) &= \frac{-3\beta^2 r^2 + 8\beta\underline{\theta} r + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta} - \underline{\theta}^2)}{2(\bar{\theta} - \underline{\theta})^2}. \end{aligned}$$

Evaluating the derivatives at $r = \underline{\theta}/\beta$ yields

$$\begin{aligned} R'_{I1}\left(\frac{\underline{\theta}}{\beta}\right) &= \frac{-2\underline{\theta} + 2\bar{\theta}}{2(\bar{\theta} - \underline{\theta})} = 1, \\ R'_{GL1}\left(\frac{\underline{\theta}}{\beta}\right) &= \frac{-3\underline{\theta}^2 + 8\underline{\theta}^2 + 2\bar{\theta}^2 - 4\underline{\theta}\bar{\theta} - 2\underline{\theta}^2}{2(\bar{\theta} - \underline{\theta})^2} = \frac{3\underline{\theta}^2 + 2\bar{\theta}^2 - 4\underline{\theta}\bar{\theta}}{2(\bar{\theta} - \underline{\theta})^2}. \end{aligned}$$

It follows that $R'_{GL1}(\underline{\theta}/\beta) > R'_{I1}(\underline{\theta}/\beta)$ since

$$\begin{aligned} R'_{GL1}\left(\frac{\underline{\theta}}{\beta}\right) &> R'_{I1}\left(\frac{\underline{\theta}}{\beta}\right) \\ \frac{3\underline{\theta}^2 + 2\bar{\theta}^2 - 4\underline{\theta}\bar{\theta}}{2(\bar{\theta} - \underline{\theta})^2} &> 1 \\ 3\underline{\theta}^2 + 2\bar{\theta}^2 - 4\underline{\theta}\bar{\theta} &> 2(\bar{\theta} - \underline{\theta})^2 = 2\bar{\theta}^2 - 4\underline{\theta}\bar{\theta} + 2\underline{\theta}^2 \\ \underline{\theta}^2 &> 0. \end{aligned}$$

That is, $R_{GL1}(r)$ intersects $R_{I1}(r)$ from below at $r = \underline{\theta}/\beta$. Since

$$R_{GL1}(0) = \frac{\underline{\theta}^3}{2\beta(\bar{\theta} - \underline{\theta})^2} > 0 > -\frac{\underline{\theta}^2}{2\beta(\bar{\theta} - \underline{\theta})} = R_{I1}(0),$$

there is an intersection of $R_{GL1}(r)$ and $R_{I1}(r)$ at some r between 0 and $\underline{\theta}/\beta$. Furthermore, we have

$$\begin{aligned} R_{I1}\left(\frac{\bar{\theta}}{\beta}\right) &= \frac{1}{\beta}\frac{\bar{\theta} + \underline{\theta}}{2}, \\ R_{GL1}\left(\frac{\bar{\theta}}{\beta}\right) &= \frac{-\frac{\bar{\theta}^3}{\beta} + 4\frac{\underline{\theta}\bar{\theta}^2}{\beta} + 2\frac{\bar{\theta}^3}{\beta} - 4\frac{\underline{\theta}\bar{\theta}^2}{\beta} - 2\frac{\underline{\theta}^2\bar{\theta}}{\beta} + \frac{\underline{\theta}^3}{\beta}}{2(\bar{\theta} - \underline{\theta})^2} = \frac{1}{\beta}\frac{\bar{\theta}^3 + \underline{\theta}^3 - 2\underline{\theta}^2\bar{\theta}}{2(\bar{\theta} - \underline{\theta})^2} = \frac{1}{\beta}\frac{\bar{\theta}^2 - \underline{\theta}^2 + \underline{\theta}\bar{\theta}}{2(\bar{\theta} - \underline{\theta})}. \end{aligned}$$

So $R_{GL1}(\bar{\theta}/\beta) > R_{I1}(\bar{\theta}/\beta)$ since

$$\begin{aligned} R_{GL1}\left(\frac{\bar{\theta}}{\beta}\right) &> R_{I1}\left(\frac{\bar{\theta}}{\beta}\right) \\ \frac{1}{\beta}\frac{\bar{\theta}^2 - \underline{\theta}^2 + \underline{\theta}\bar{\theta}}{2(\bar{\theta} - \underline{\theta})} &> \frac{1}{\beta}\frac{\bar{\theta}^2 - \underline{\theta}^2}{2(\bar{\theta} - \underline{\theta})} \\ \bar{\theta}^2 - \underline{\theta}^2 + \underline{\theta}\bar{\theta} &> \bar{\theta}^2 - \underline{\theta}^2 \\ \underline{\theta}\bar{\theta} &> 0. \end{aligned}$$

Since $R_{GL1}(r) < R_{I1}(r)$ as r grows large, there is an intersection of $R_{GL1}(r)$ and $R_{I1}(r)$ at some $r > \bar{\theta}/\beta$. So we have identified three points of intersection of $R_{I1}(r)$ and $R_{GL1}(r)$. Since $R_{I1}(r)$ and $R_{GL1}(r)$ are second-order and third-order polynomials, respectively, there cannot be any further intersections, so that

$$R_{GL1}(r) > R_{I1}(r), \quad \frac{\underline{\theta}}{\beta} < r \leq \frac{\bar{\theta}}{\beta}.$$

From the definition of case L (i.e., $\underline{\theta}/\beta \leq r \leq \bar{\theta}/(2\beta)$), it follows that except at the lower boundary $r = \underline{\theta}/\beta$, we have

$$R_G(r) > R_I(r), \quad \text{case L.}$$

An example is illustrated in Figure 3.18.

b) Case H

Let $R_{GH1}(r)$ denote the function on the the right-hand side of (3.12) for $\alpha = 1$:

$$R_{GH1}(r) = \frac{\beta^2 r^3 - \beta\bar{\theta}3r^2 + \bar{\theta}^2 3r - \frac{1}{\beta}(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta} - \underline{\theta})^2}.$$

Evaluating $R_{GH1}(r)$ at $r = \underline{\theta}/\beta$ yields

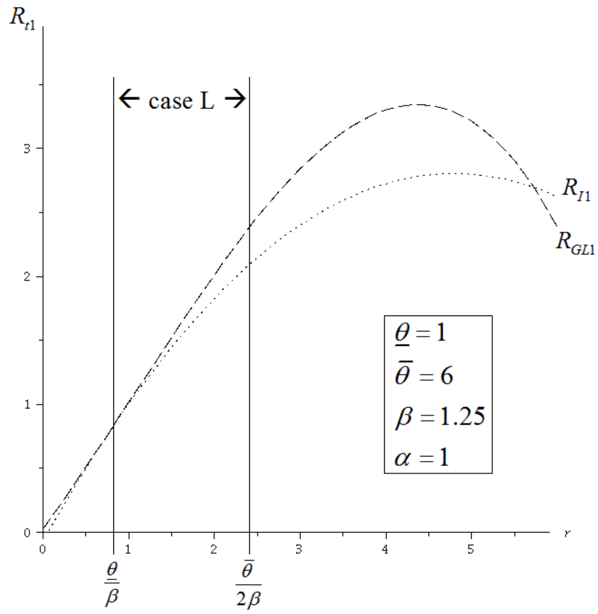


Figure 3.18: Exemplary expected repayment functions for $\alpha = 1$, case L.

$$\begin{aligned} R_{GH1}\left(\frac{\underline{\theta}}{\beta}\right) &= \frac{\beta^2\left(\frac{\underline{\theta}}{\beta}\right)^3 - 3\beta\bar{\theta}\left(\frac{\underline{\theta}}{\beta}\right)^2 + 3\bar{\theta}^2\frac{\underline{\theta}}{\beta} - \frac{1}{\beta}(\underline{\theta}^2\bar{\theta} + \underline{\theta}\bar{\theta}^2 - \underline{\theta}^3)}{2(\bar{\theta} - \underline{\theta})^2} \\ &= \frac{\underline{\theta}}{\beta}\frac{\underline{\theta}^2 - 3\underline{\theta}\bar{\theta} + 3\bar{\theta}^2 - \underline{\theta}\bar{\theta} - \bar{\theta}^2 + \underline{\theta}^2}{2(\bar{\theta} - \underline{\theta})^2} = \frac{\underline{\theta}}{\beta}\frac{2\underline{\theta}^2 + 2\bar{\theta}^2 - 4\underline{\theta}\bar{\theta}}{2(\bar{\theta} - \underline{\theta})^2} = \frac{\underline{\theta}}{\beta}, \end{aligned}$$

which equals $R_{I1}\left(\frac{\underline{\theta}}{\beta}\right)$. We have seen in the main text (and in the second part of Appendix 3.8.3) that

$$R_{GH1}\left(\frac{\bar{\theta}}{\beta}\right) = R_{I1}\left(\frac{\bar{\theta}}{\beta}\right) = \frac{1}{\beta}\frac{\underline{\theta} + \bar{\theta}}{2}.$$

Differentiating $R_{GH1}(r)$ with respect to r gives

$$R'_{GH1}(r) = \frac{3\beta^2 r^2 - 6\beta\bar{\theta}r + 3\bar{\theta}^2}{2(\bar{\theta} - \underline{\theta})^2} = \frac{3\beta^2\left(r - \frac{\bar{\theta}}{\beta}\right)^2}{2(\bar{\theta} - \underline{\theta})^2}.$$

Evaluating $R'_{I1}(r)$ and $R'_{GH1}(r)$ at $r = \bar{\theta}/\beta$ gives

$$R'_{I1}\left(\frac{\bar{\theta}}{\beta}\right) = R'_{GH1}\left(\frac{\bar{\theta}}{\beta}\right) = 0.$$

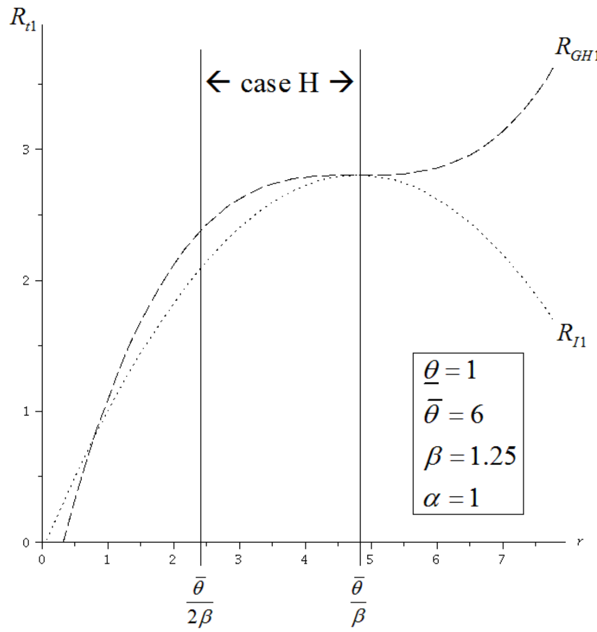


Figure 3.19: Exemplary expected repayment functions for $\alpha = 1$, case H.

So $R_{I1}(r)$ and $R_{GH1}(r)$ have one intersection at $r = \underline{\theta}/\beta$ and a 'double intersection' at $r = \bar{\theta}/\beta$. Given that $R_{I1}(r)$ and $R_{GH1}(r)$ are second-order and third-order polynomials, respectively, there are no further intersections. It follows that

$$R_{GH1}(r) > R_{I1}(r), \quad \frac{\underline{\theta}}{\beta} < r < \frac{\bar{\theta}}{\beta}.$$

So, except at $r = \bar{\theta}/\beta$, we have

$$R_G(r) > R_I(r), \quad \text{case H.}$$

An example is illustrated in Figure 3.19. Taken together, we have proven that $R_G(r) > R_I(r)$ for $\frac{\underline{\theta}}{\beta} < r < \frac{\bar{\theta}}{\beta}$ when $\alpha = 1$.

3.8.7 Proof: $U_I(r) > U_G(r)$ for cases L and H.

We abstain from an algebraic proof, but provide a verbal proof instead. This proof is more intuitive. It is based on Figure 3.3 and we do it for case L. It will become clear that this also proves case H. We compare the expected utility of two IL borrowers to the expected utility of two borrowers with a GL contract. The proof is done for the same r so that the payoff distribution in Figure 3.3 applies to the joint liability group and to both IL borrowers.

First, joint liability does not affect utility in cases (AA) and (BB) since payoffs per borrower are identical for the group and the two IL borrowers. In case (BC), one of the group members repays $2r$ and the other free-rides. Each of the two IL borrowers repays r. Since borrowers are risk-neutral and the game is symmetric, expected utility is the same for the group and the IL borrowers. The latter thought also applies to case (CC) (cf. Assumption 4). Thus, we are left with cases (AB) and (AC).

Consider case (AB). The two IL borrowers have payoffs $\theta_1 - p(\theta_1)$ and $\theta_2 - r$ (if $\theta_1 \in A$ and $\theta_2 \in B$). Due to symmetry, expected utility in case (AB) for an IL borrower is $\frac{1}{2}(\theta_1 - p(\theta_1) + \theta_2 - r)$. By contrast, the group has payoffs $\theta_1 - p(\theta_1)$ and $\theta_2 - p(\theta_2)$, and expected utility $\frac{1}{2}(\theta_1 - p(\theta_1) + \theta_2 - p(\theta_2))$ per borrower. Since $p(\theta_2) > r$, non-cooperative behavior and the inability to make partial repayments cause the expected utility of a borrower in a group to be lower.

Thus, we have already proven case H. To complete the proof for case L, consider case (AC). The two IL borrowers have payoffs $\theta_1 - p(\theta_1)$ and $\theta_2 - r$ (if $\theta_1 \in A$ and $\theta_2 \in C$), so that expected utility in (AC) for an IL borrower is $\frac{1}{2}(\theta_1 - p(\theta_1) + \theta_2 - r)$. By contrast, the group has payoffs θ_1 and $\theta_2 - 2r$, and expected utility $\frac{1}{2}(\theta_1 + \theta_2 - 2r)$. Since $p(\theta_1) < r$, the expected utility of a borrower in a group is lower.

3.8.8 Proof of Proposition 3.3

First, we have to show that $\frac{\bar{\theta}^2}{4\beta(\bar{\theta}-\underline{\theta})}$ is indeed the maximum under IL when $\alpha = 0$. Define the right-hand side of (3.8) for $\alpha = 0$, expected repayment under IL when penalties are non-pecuniary, as $R_{I0}(r)$, and its derivative with respect to r as $R'_{I0}(r)$:

$$\begin{aligned} R_{I0}(r) &= \frac{-2\beta r^2 + 2\bar{\theta} r}{2(\bar{\theta} - \underline{\theta})}, \\ R'_{I0}(r) &= \frac{-4\beta r + 2\bar{\theta}}{2(\bar{\theta} - \underline{\theta})}. \end{aligned}$$

The loan rate $r = \bar{\theta}/(2\beta)$ maximizes $R_{I0}(r)$. The maximum value is

$$R_{I0}\left(\frac{\bar{\theta}}{2\beta}\right) = \frac{-2\beta\left(\frac{\bar{\theta}}{2\beta}\right)^2 + 2\bar{\theta}\left(\frac{\bar{\theta}}{2\beta}\right)}{2(\bar{\theta}-\underline{\theta})} = \frac{\bar{\theta}^2}{4\beta(\bar{\theta}-\underline{\theta})}.$$

To prove Proposition 3.3, it remains to show that maximum expected repayment under GL is less than that for cases L and H.

a) Case L

Let $R_{GL0}(r)$ denote the right-hand side of (3.10) for $\alpha = 0$:

$$R_{GL0}(r) = \frac{-6\beta^2 r^3 + 8\beta\underline{\theta}r^2 + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta})r}{2(\bar{\theta}-\underline{\theta})^2}. \tag{3.32}$$

We have to show that

$$R_{GL0}(r) < \frac{\bar{\theta}^2}{4\beta(\bar{\theta}-\underline{\theta})}.$$

Notice that

$$R_{I0}(0) = R_{GL0}(0) = 0.$$

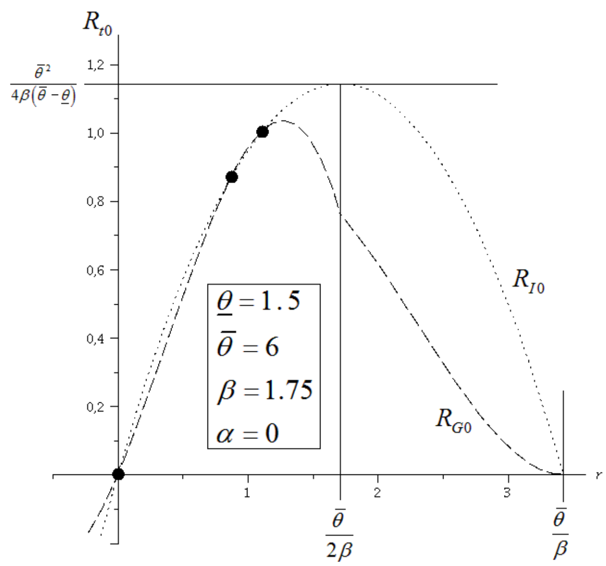
We have $R_{GL0}(r) < R_{I0}(r)$ iff

$$\begin{aligned}
\frac{-6\beta^2 r^3 + 8\beta\underline{\theta}r^2 + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta})r}{2(\bar{\theta}-\underline{\theta})^2} &< \frac{-2\beta r^2 + 2\bar{\theta}r}{2(\bar{\theta}-\underline{\theta})} \\
-6\beta^2 r^2 + 8\beta\underline{\theta}r + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta}) &< (-2\beta r + 2\bar{\theta})(\bar{\theta}-\underline{\theta}) = -2\beta(\bar{\theta}-\underline{\theta})r + 2(\bar{\theta}^2 - \underline{\theta}\bar{\theta}) \\
-6\beta^2 r^2 + 8\beta\underline{\theta}r - 2\underline{\theta}\bar{\theta} &< -2\beta(\bar{\theta}-\underline{\theta})r \\
6\beta^2 r^2 - 2\beta(\bar{\theta}+3\underline{\theta})r + 2\underline{\theta}\bar{\theta} &> 0 \\
r^2 - \frac{\bar{\theta}+3\underline{\theta}}{3\beta}r + \frac{\underline{\theta}\bar{\theta}}{3\beta^2} &> 0.
\end{aligned} \tag{3.33}$$

The roots of the polynomial on the left-hand side are:

$$
\begin{aligned}
r_{1/2} &= \frac{\bar\theta + 3\underline{\theta}}{6\beta} \pm \sqrt{\frac{1}{4}\left(\frac{\bar\theta + 3\underline{\theta}}{3\beta}\right)^2 - \frac{\underline{\theta}\bar\theta}{3\beta^2}} = \frac{\bar\theta + 3\underline{\theta}}{6\beta} \pm \frac{1}{6\beta}\sqrt{\bar\theta^2 - 6\underline{\theta}\bar\theta + 9\underline{\theta}^2} \\
&= \frac{\bar\theta + 3\underline{\theta}}{6\beta} \pm \frac{1}{6\beta}\sqrt{(\bar\theta - 3\underline{\theta})^2} = \frac{(\bar\theta + 3\underline{\theta}) \pm (\bar\theta - 3\underline{\theta})}{6\beta} = \left\{\frac{\underline{\theta}}{\beta}, \frac{\bar\theta}{3\beta}\right\} < \frac{\bar\theta}{2\beta}.
\end{aligned}
$$

So the three intersections of $R_{GL0}(r)$ and $R_{I0}(r)$ occur at $r = 0$, $r = r_1 = \underline{\theta}/\beta$, and $r = r_2 = \bar\theta/(3\beta)$. An example can be seen in the following graph:



Suppose first that $\bar\theta \leq 3\underline{\theta}$, so that

$$r_2 = \frac{\bar\theta}{3\beta} < \frac{\underline{\theta}}{\beta} = r_1.$$

Then

$$R_{GL0}(r) < R_{I0}(r), \quad r > \frac{\underline{\theta}}{\beta},$$

since $\frac{\underline{\theta}}{\beta}$ is the largest intersection and $R_{GL0}(r) < R_{I0}(r)$ for large r (cf. inequality (3.33)). Next, consider the case $\bar\theta > 3\underline{\theta}$, in which $r_2 = \bar\theta/(3\beta) > \underline{\theta}/\beta = r_1$. Differentiating $R_{GL0}(r)$ with respect to r gives

$$R'_{GL0}(r) = \frac{-18\beta^2 r^2 + 16\beta\underline{\theta}r + 2(\bar\theta^2 - 2\underline{\theta}\bar\theta)}{2(\bar\theta - \underline{\theta})^2}. \tag{3.34}$$

So

$$\begin{aligned} R'_{GL0}(0) &= \frac{\bar{\theta}(\bar{\theta}-2\underline{\theta})}{(\bar{\theta}-\underline{\theta})^2} > 0, \\ R'_{GL0}\left(\frac{\bar{\theta}}{3\beta}\right) &= \frac{-18\beta^2\left(\frac{\bar{\theta}}{3\beta}\right)^2 + 16\beta\underline{\theta}\frac{\bar{\theta}}{3\beta} + 2(\bar{\theta}^2 - 2\underline{\theta}\bar{\theta})}{2(\bar{\theta}-\underline{\theta})^2} \\ &= \frac{-2\bar{\theta}^2 + \frac{16}{3}\bar{\theta}\underline{\theta} + 2\bar{\theta}^2 - 4\bar{\theta}\underline{\theta}}{2(\bar{\theta}-\underline{\theta})^2} = \frac{2\bar{\theta}\underline{\theta}}{3(\bar{\theta}-\underline{\theta})^2} > 0. \end{aligned}$$

The fact that $R_{GL0}(r)$, a cubic polynomial with negative leading coefficient (viz., $\frac{-6\beta^2}{2(\bar{\theta}-\underline{\theta})^2}$), is upward-sloping at $r = 0$ and at $r = \bar{\theta}/(3\beta)$ implies that it is upward-sloping in between. It follows that

$$R_{GL0}(r) < R_{GL0}\left(\frac{\bar{\theta}}{3\beta}\right) = R_I 0\left(\frac{\bar{\theta}}{3\beta}\right), \quad r < \frac{\bar{\theta}}{3\beta},$$

and, since $\frac{\theta}{\beta}$ is the largest intersection and $R_{GL0}(r) < R_{I0}(r)$ for large r (cf. inequality (3.33)),

$$R_{GL0}(r) < R_{I0}(r), \quad r > \frac{\bar{\theta}}{3\beta}.$$

This completes the proof that

$$R_{GL0}(r) < \frac{\bar{\theta}^2}{4\beta(\bar{\theta}-\underline{\theta})}, \quad \text{case L.}$$

b) Case H

Let $R_{GH0}(r)$ denote the right-hand side of (3.12) for $\alpha = 0$ and $R'_{GH0}(r)$ its derivative with respect to r:

$$\begin{aligned} R_{GH0}(r) &= \frac{2\beta^2 r^3 - 4\beta\bar{\theta}r^2 + 2\bar{\theta}^2 r}{2(\bar{\theta}-\underline{\theta})^2}, \\ R'_{GH0}(r) &= \frac{6\beta^2 r^2 - 8\beta\bar{\theta}r + 2\bar{\theta}^2}{2(\bar{\theta}-\underline{\theta})^2}. \end{aligned}$$

We have $R'_{GH0}(r) = 0$ iff

$$
\begin{aligned}
0 &= 6\beta^2 r^2 - 8\beta\bar{\theta}r + 2\bar{\theta}^2 \\
0 &= r^2 - \frac{4}{3}\frac{\bar{\theta}}{\beta}r + \frac{1}{3}\left(\frac{\bar{\theta}}{\beta}\right)^2 \\
r_{1/2} &= \frac{2}{3}\frac{\bar{\theta}}{\beta} \pm \sqrt{\frac{4}{9}\left(\frac{\bar{\theta}}{\beta}\right)^2 - \frac{1}{3}\left(\frac{\bar{\theta}}{\beta}\right)^2} = \frac{2}{3}\frac{\bar{\theta}}{\beta} \pm \frac{1}{3}\frac{\bar{\theta}}{\beta} = \left\{\frac{\bar{\theta}}{3\beta}, \frac{\bar{\theta}}{\beta}\right\}.
\end{aligned}
$$

$R_{GH0}(r)$ is downward-sloping in the interval $(\bar{\theta}/(3\beta), \bar{\theta}/\beta)$ since it is a cubic polynomial with positive leading coefficient (viz., $\frac{\beta^2}{(\bar{\theta}-\underline{\theta})^2}$) and extrema at $\frac{\bar{\theta}}{3\beta}$ and $\frac{\bar{\theta}}{\beta}$. Since $\frac{\bar{\theta}}{3\beta} < \frac{\bar{\theta}}{2\beta} < \frac{\bar{\theta}}{\beta}$,

$$
R_{GH0}(r) \leq R_{GH0}\left(\frac{\bar{\theta}}{2\beta}\right),
$$

From continuity of $R_G(r)$ between cases L and H (cf. Appendix 3.8.2),

$$
R_{GH0}\left(\frac{\bar{\theta}}{2\beta}\right) = R_{GL0}\left(\frac{\bar{\theta}}{2\beta}\right).
$$

Using the fact that, in case L,

$$
R_{GL0}\left(\frac{\bar{\theta}}{2\beta}\right) < \frac{\bar{\theta}^2}{4\beta(\bar{\theta}-\underline{\theta})},
$$

it follows that

$$
R_{GH0}(r) < \frac{\bar{\theta}^2}{4\beta(\bar{\theta}-\underline{\theta})}, \quad \text{case H.}
$$

This completes the proof.

3.8.9 Proof that $R_C(r)$ is hump-shaped, and $R_C(r) \geq R_I(r)$

The hump-shape occurs over the interval defined in (3.16). Differentiating the expression for $R_C(r)$ in (3.15) yields

$$
R'_C(r) = \frac{-6\beta^2 r^2 + 8\beta\underline{\theta}r + (\bar{\theta}^2 - 2\bar{\theta}\underline{\theta} - \underline{\theta}^2)}{(\bar{\theta}-\underline{\theta})^2}.
$$

So

$$R'_C\left(\frac{\underline{\theta}}{\beta}\right) = 1 > 0, \quad R'_C\left(\frac{\bar{\theta}+\underline{\theta}}{2\beta}\right) = -\frac{(\bar{\theta}^2-\underline{\theta}^2)+2\underline{\theta}(\bar{\theta}-\underline{\theta})}{2(\bar{\theta}-\underline{\theta})^2} < 0.$$

As $R_C(r)$ is a cubic polynomial with negative leading coefficient, it follows that it is hump-shaped over the interval $[\underline{\theta}/\beta, (\bar{\theta}+\underline{\theta})/(2\beta)]$. Note that $\underline{\theta}/\beta < (\bar{\theta}+\underline{\theta})/(2\beta)$ since $\frac{\bar{\theta}}{2} > \underline{\theta}$.

Proof that $R_C(r) \geq R_I(r)$ when (3.16) is satisfied:

From (3.8) and (3.15), $R_C(r) \geq R_I(r)$ exactly if

$$r^2 - \frac{\bar{\theta}+3\underline{\theta}}{2\beta}r + \frac{\underline{\theta}(\bar{\theta}+\underline{\theta})}{2\beta^2} \leq 0.$$

The roots of the quadratic equation on the left-hand side are $r = \underline{\theta}/\beta$ and $r = (\bar{\theta}+\underline{\theta})/(2\beta)$. So $R_C(r) \geq R_I(r)$ between these loan rates.

3.8.10 Proof: Proposition 3.6 for large r

It suffices to show that with GL and $r > (\bar{\theta}+\underline{\theta})/(2\beta)$, the expected repayment falls short of R_C^{max} and the deadweight loss is an increasing function of the loan rate. For $r > (\bar{\theta}+\underline{\theta})/(2\beta)$, cases (AC), (BC), and (CC) drop out, and the repayment function becomes

$$\tilde{R}_C(r) = \frac{2\beta^2r^3 - 4\beta\bar{\theta}r^2 + 2\bar{\theta}^2r}{(\bar{\theta}-\underline{\theta})^2}.$$

Using (3.15), we have

$$\tilde{R}_C\left(\frac{\bar{\theta}+\underline{\theta}}{2\beta}\right) = \frac{\bar{\theta}+\underline{\theta}}{4\beta} = R_C\left(\frac{\bar{\theta}+\underline{\theta}}{2\beta}\right),$$

so the expected repayment function with GL is continuous. Differentiating $\tilde{R}_C(r)$ gives

$$\tilde{R}'_C(r) = \frac{6\beta^2r^2 - 8\beta\bar{\theta}r + 2\bar{\theta}^2}{(\bar{\theta}-\underline{\theta})^2}.$$

The roots of the quadratic equation on the right-hand side are $\bar{\theta}/(3\beta)$ and $\bar{\theta}/\beta$. $\tilde{R}_C(r)$ takes on its local maximum and minimum, respectively, at these loan rates. It follows that

$\tilde{R}_C(r)$ is downward-sloping between $(\bar{\theta}+\underline{\theta})/(2\beta)$ and $\bar{\theta}/\beta$. Taken together, it follows that $\tilde{R}_C(r) < R_C^{max}$ for $(\bar{\theta}+\underline{\theta})/(2\beta) < r < \bar{\theta}/\beta$. The deadweight loss is

$$\tilde{D}_C(r) = \frac{-8\beta^3 r^3 + 12\beta^2\bar{\theta}r^2 - \bar{\theta}^3 - 3\bar{\theta}^2\underline{\theta} - 3\bar{\theta}\underline{\theta}^2 + 3\underline{\theta}^3}{6\beta(\bar{\theta}-\underline{\theta})^2}$$

for $r > (\bar{\theta}+\underline{\theta})/(2\beta)$. The deadweight loss is a continuous function of the loan rate:

$$\tilde{D}_C\left(\frac{\bar{\theta}+\underline{\theta}}{2\beta}\right) = \frac{\bar{\theta}^3 - 3\bar{\theta}\underline{\theta}^2 + 2\underline{\theta}^3}{6\beta(\bar{\theta}-\underline{\theta})^2} = D_C\left(\frac{\bar{\theta}+\underline{\theta}}{2\beta}\right).$$

It increases as the loan rate rises:

$$\tilde{D}_C'(r) = \frac{4\beta r(\bar{\theta}-\beta r)}{(\bar{\theta}-\underline{\theta})^2} > 0$$

for all $r < \bar{\theta}/\beta$.

3.8.11 Proof that $R_S(r)$ is hump-shaped with social sanctions and $\alpha = 0$

Using equation (3.20), expected repayment is

$$R_S(r) = \Pi_S(r)r = \frac{-\beta^2 r^3 + 2\beta\underline{\theta}r^2 + (\bar{\theta}^2 - 2\underline{\theta}\bar{\theta})r}{(\bar{\theta}-\underline{\theta})^2}.$$

Its derivative with respect to r is

$$R_S'(r) = \frac{-3\beta^2 r^2 + 4\beta\underline{\theta}r + \bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{(\bar{\theta}-\underline{\theta})^2}.$$

At $r = \frac{\underline{\theta}}{\beta}$, expected repayment becomes

$$R_S\left(\frac{\underline{\theta}}{\beta}\right) = \frac{-\frac{\underline{\theta}^3}{\beta} + 2\frac{\underline{\theta}^3}{\beta} + \frac{\underline{\theta}\bar{\theta}^2}{\beta} - 2\frac{\underline{\theta}^2\bar{\theta}}{\beta}}{(\bar{\theta}-\underline{\theta})^2} = \frac{\underline{\theta}}{\beta(\bar{\theta}-\underline{\theta})^2}(\underline{\theta}^2 + \bar{\theta}^2 - 2\bar{\theta}\underline{\theta}) = \frac{\underline{\theta}}{\beta(\bar{\theta}-\underline{\theta})^2}(\bar{\theta}-\underline{\theta})^2 = \frac{\underline{\theta}}{\beta} > 0.$$

The derivative at $r = \frac{\underline{\theta}}{\beta}$ is

$$R_S'\left(\frac{\underline{\theta}}{\beta}\right) = \frac{-3\beta^2\left(\frac{\underline{\theta}}{\beta}\right)^2 + 4\beta\underline{\theta}\frac{\underline{\theta}}{\beta} + \bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{(\bar{\theta}-\underline{\theta})^2} = \frac{-3\underline{\theta}^2 + 4\underline{\theta}^2 + \bar{\theta}^2 - 2\underline{\theta}\bar{\theta}}{(\bar{\theta}-\underline{\theta})^2} = \frac{(\bar{\theta}-\underline{\theta})^2}{(\bar{\theta}-\underline{\theta})^2} = 1 > 0.$$

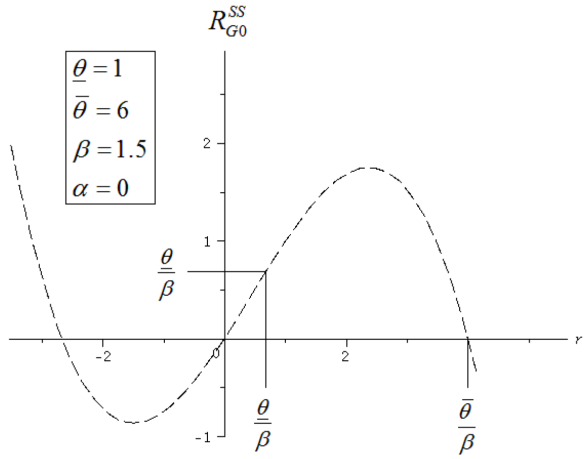


Figure 3.20: Expected repayments under GL with social sanctions and $\alpha = 0$.

Furthermore, expected repayment at $r = \frac{\bar{\theta}}{\beta}$ is

$$R_S\left(\frac{\bar{\theta}}{\beta}\right) = \frac{-\beta^2 \frac{\bar{\theta}^3}{\beta^3} + 2\beta\underline{\theta}\frac{\bar{\theta}^2}{\beta^2} + (\bar{\theta}^2 - 2\bar{\theta}\underline{\theta})\frac{\bar{\theta}}{\beta}}{(\bar{\theta} - \underline{\theta})^2} = \frac{\frac{\bar{\theta}^3}{\beta} + \frac{2\underline{\theta}\bar{\theta}^2}{\beta} - \frac{\bar{\theta}^3}{\beta} - \frac{2\underline{\theta}\bar{\theta}^2}{\beta}}{(\bar{\theta} - \underline{\theta})^2} = 0.$$

Since $R_S(r)$ is a cubic polynomial with negative leading coefficient, the fact that it is positive and upward sloping at $r = \frac{\underline{\theta}}{\beta}$ and zero at $r = \frac{\bar{\theta}}{\beta}$ implies that it has no root in the interval $(\frac{\underline{\theta}}{\beta}, \frac{\bar{\theta}}{\beta})$. That is, expected repayment is positive and takes on a unique local maximum in the interval. In other words, expected repayments under GL with social sanctions and $\alpha = 0$ are hump-shaped. Figure 3.20 illustrates this.

Chapter 4

Portfolio Choice with Social Returns

This chapter is based on joint work with Michaela Leidl and Gregor Dorfleitner. It contains elements of Dorfleitner, Leidl, and Reeder (2009).

4.1 Motivation and the literature

The financial crisis has led many to question their investments. Apart from rethinking the risks involved, non-financial objectives have gained in importance. The terms ‘Social Business’ and ‘Socially Responsible Investing’ (SRI) are currently under discussion both in the general public and between researchers of various fields.[1] Even though indisputable definitions are missing, the bottom line seems clear: Many people care about more than just financial returns. From a practical point of view, many of these people have too limited a circle of influence to actively engage in fostering social business. Apart from voluntary work in their leisure time, these people might express their preferences by taking into account the social dimension in their investment decisions. This tendency has brought up ‘social funds’, e.g., microfinance investment funds (MFIFs), renewable energy funds, and the like. The aim of these funds is to combine social and financial returns to attract investors.

Among other things, the findings of Markowitz (1952) prompted the implementation of mutual funds. His work is surely one of the most influential papers in the finance literature. However, Markowitz’ setup exclusively relies on financial returns. Is it possible to build a theory which explains the existence and the composition of social funds not by resorting to irrational social investors, but by assuming rational optimization of investors? This is the aim of the current chapter.

In doing so, we strongly build on Markowitz (1952). Apart from being a very ambitious task, it would be well beyond the scope of this chapter to enumerate and classify all research based on Markowitz’ work in this literature survey,[2] especially since our model directly extends his work. Nevertheless, apart from presenting his main results, we want to give a sketch of the finance literature following his early work since we hope to be able to extend our model in similar ways in the future, for instance to apply it to the realm of asset pricing.

In his seminal work ‘Portfolio Selection’ Markowitz (1952) recognized the crucial role of risk and proposed to use the statistical concept of an asset’s variance as an appropriate measure for risk. This measure is used until today, although other measures have been proposed, e.g., ‘value at risk’ (mainly in judging risks of credit portfolios) or ‘downside risk’. The

[1] A good starting point into the literature on SRI is the categorization of Hoepner (2007). The term ‘social business’ is strongly coined by Nobel Peace Prize winner Muhammad Yunus, e.g., see Yunus (2007). Yunus is also engaged in the yearly summit on social business organized by the Berlin-based Genisis institute.

[2] As of 05.12.2009, Google Scholar shows 8500 papers which quote Markowitz (1952).

latter concept is theoretically appealing since it separates downside risk and upside potential. Clearly, it is only downside variations which investors dislike. Grootveld and Hallerbach (1999) compare different measures of downside risk to the variance, both theoretically and based on their implications for optimal asset allocation. In terms of theory, they show that the performance of different measures for downside risk varies considerably when compared to the variance, and that only a few of those measures perform strictly better. Against the background of their findings, it is no big surprise that the variance is still a commonly used risk measure, and we use it in this chapter, too.

Another key insight of Markowitz concerns the effects of diversification. Dependencies of returns of individual assets among each other crucially influence the volatility of a portfolio built from these assets. Thus, an appropriate choice of assets can reduce a portfolio's risk without diminishing expected returns. Moreover, Markowitz' suggestions allowed to formulate the decision making process of portfolio choice as an optimization problem. He suggested that investors should[3] maximize the expected return of a portfolio and minimize its variance. The implied behavior of investors can be seen as an alternative approach to decision making under uncertainty, i.e., as an alternative (in fact, it is a special case) to the expected utility criterion proposed by von Neumann and Morgenstern (1944).

A result directly related to Markowitz' findings is the concept of 'two fund separation', sometimes called 'mutual fund separation', or 'Tobin separation' alluding to James Tobin (1958) who first proposed the idea. We will have much to say about this phenomenon in the remainder of this chapter. In fact, two of our main results are directly related to Tobin's findings. Tobin analyzed investors that face a set of risky assets as well as a riskless asset and decide about how to allocate funds using a mean-variance objective function. He showed that the optimal portfolio of risky assets has the same composition for all investors, irrespective of their risk aversion. Thus, differences in risk aversion among investors only lead to different shares of wealth invested in that same portfolio, the remaining share of wealth being invested in the riskless asset.

Based on Tobin's findings, a natural question arose: If all investors choose the same portfolio of risky assets, can this information about the demand for assets be used to determine

[3] The original aim of Markowitz was to provide investors with an investment recipe and, thus, rather normative.

asset prices? The answer to that question is the famous capital asset pricing model (CAPM), which is mainly ascribed to Sharpe (1964), Lintner (1965), and Mossin (1966). These authors suggested to ask for conditions that must be met so that the optimal, tangent portfolio is an equilibrium. From these conditions they derive that the crucial determinant of asset prices is the expected return of the market and the comovement of an asset with the market. We abstain from a discussion on the advantages and inconveniences of the model here, but refer the reader to Chapter 8 in Cochrane (2005) for a detailed analysis in an advanced formal setup.

The idea of the CAPM was extended by Merton (1971, 1973) to a continuous time setup with dynamic portfolio choice, called the intertemporal capital asset pricing model (ICAPM). Arbitrage pricing theory (APT) as first proposed by Ross (1976) belongs to the class of multi-factor pricing models. While the CAPM determines asset prices by their relative performance compared to just one factor, viz., the market, APT allows to include several (e.g., macro-economic) factors to account for systematic risk. In fact, Ross, Westerfield, and Jaffe (2005, p.309-10) claim that the CAPM can be treated as a one-factor special case of the APT in terms of its implications, although both approaches differ considerably in terms of their origin and application. A multi-factor model which has received much attention in the literature is the 'three-factor model' by Fama and French (1995) which adds the size of the firm underlying an asset and the book-to-market ratio to the market variable in order to predict excess returns. They find an empirically significant influence for all three variables, which contradicts the CAPM predictions. However, empirical tests are not conclusive neither in accepting nor rejecting the CAPM.

Another approach to asset pricing known as the consumption capital asset pricing model (CCAPM) is ascribed to Lucas (1978) and Breeden (1979). In analogy to the CAPM, it uses an indicator of risk to determine the excess return of an asset. In contrast to the CAPM, which measures risk as the covariance of an asset's return with the market (the 'market β'), the CCAPM measures risk as the comovement of an asset's returns with consumption (the 'consumption β').

All these models are based on the assumption that the price of an asset is determined by its financial characteristics. The ultimate aim of our research is to extend standard portfolio theory in order to propose a theory of asset pricing similar to the CAPM. As Luenberger

(1998, p.222) notes, "Markowitz and the CAPM are beautiful theories that ushered in an era of quantitative analysis and have provided an elegant foundation to support further work". In this chapter and the corresponding paper, we extend standard mean-variance theory by adding a social dimension.

Clearly, we are not the first to suggest the use of social returns. The most common method to include a social dimension into investment choice is screening.[4] The idea is simple: From all available assets, investors choose the subset of assets they are willing to invest in. 'Positive screening' picks out the assets to invest in, whereas 'negative screening' excludes assets which the investor does not want to fund under any circumstances. Screening usually takes place prior to any kind of optimization (if there is one).[5] The optimization after screening could be done à la Markowitz, i.e., only depending on financial characteristics. In that case, once the subset of acceptable assets is chosen, there is no quantification of social returns. The combined procedure of screening and standard portfolio optimization might be considered a bounded rationality approach in that investors optimize only after having greatly reduced the complexity of the problem by having limited their choices available. Dupré, Girerd-Potin, and Kassoua (2004) apply screening to a large set of assets, determine the pre- and post-screening efficient frontiers and confirm intuition: By reduced diversification possibilities, the efficient portfolios after screening are financially worse than the ones before screening.[6]

Social returns of individual assets differ considerably and might even be considered stochastic, as already noted by Dupré, Girerd-Potin, and Kassoua (2004). These authors point to variations of companies' social behavior over time, which we consider a valid reason to model social returns as stochastic.[7]

The starting point for both the deterministic and the stochastic analysis is a metric scale to measure social returns. Once such a metric exists, empirical estimates for statistical moments of social returns, in particular means, variances, and covariances can be derived. Thus, assets

[4]For instance, see D'Antonio and Johnsen (1997); D'Antonio, Johnsen, and Hutton (2000); Renneboog, Ter Horst, and Zhang (2008).

[5]Screening is not necessarily done before anything else. One could imagine a repeated optimization-screening pattern. After an initial optimization à la Markowitz, the optimal weights of assets could be checked and the assets (or amount of assets) incompatible with the investor's preferences could be sorted out. The remaining assets could then be subjected to another round of optimization.

[6]Technically, finding the efficient frontier after screening boils down to imposing additional constraints on a maximization problem.

[7]They mention imperfect measurability as another reason.

do no longer have to be reduced to their financial characteristics, but have both financial and social returns. Dependency structures between social returns of different assets can be estimated, as well as covariances between financial returns of one asset and social returns of another. Apart from these inter-asset dependencies, some investments can have significant intra-asset covariances: A high financial performance goes at the expense of a low social performance, and vice versa. Also, financial and social returns of an asset might depend positively on each other: A high financial performance is accompanied by a high social return.

Since uncertainty about financial returns can be accompanied by uncertainty about social returns, we consider stochastic social returns along with stochastic financial returns in our most general setup in Subsection 4.2.1. In Subsection 4.2.2, we confine the analysis to a less complicated maximization problem. This latter problem can be interpreted as representing one of two situations: In the first, social returns are deterministic. In the second, they are stochastic, but investors do not care about variations of social returns (the investors could then be called 'risk-neutral regarding social returns'). Although considerably different conceptually, both situations are mathematically similar, in fact almost identical. Another reason why we consider the restricted problem in Subsection 4.2.2 is that estimates for variations in social returns are rather unreliable due to data restrictions.[8]

Deterministic optimization problems similar to our proposal have already been proposed and applied by Dupré, Girerd-Potin, and Kassoua (2004). However, they do neither present nor interpret any of the theoretical results we derive in Subsection 4.2.2. Furthermore, the efficient frontier they discuss is the standard two-dimensional financial mean-variance concept, although portfolios are assigned three dimensions. While we define, discuss and apply the concept of the efficient frontier in a five-dimensional space in the stochastic setup, we give a more intuitive discussion of the efficient frontier for deterministic social returns in three-dimensional space. In that respect, we want to mention Dunn (2006) who took an educated guess on the shape of a three-dimensional efficient frontier. We will confirm some general properties of his proposal in the deterministic setup, but add many others.

[8]We do not work with real world data in this chapter, except when we construct the graph of the three-dimensional efficient frontier in Appendix 4.4.1. However, it is one of the main objectives of Dorfleitner, Leidl, and Reeder (2009) to bring the model of this chapter to the data.

4.2 Theory

As in standard portfolio theory, we consider a one-period investment problem, i.e., an investor at two points in time, where decisions are only taken at the first. This decision is about asset allocation: The investor decides about how much to invest in which assets. At the second point in time, two types of returns accrue: financial returns R and social returns S. Investors' preferences are formulated in terms of simple returns. This means that an investor allocating an initial wealth of v_0 gets a financial payoff of $v_0 \cdot (1 + R)$ plus a social 'payoff' of $v_0 \cdot S$ at the end of the period.[9] However, note that both payoffs are not directly transferable into each other. Financial returns are measured in terms of money per unit invested, whereas the social return is a non-monetary value per unit invested. The measurement of social returns is a problem in reality which we ignore in this chapter.[10] As in standard mean-variance analysis, using rates of return instead of absolute financial and social wealth levels implies that wealth effects have to be captured by the coefficients of the objective function to be defined further below. We assume that there is a metric for social returns which allows us to rank alternative investments in terms of social returns. This might be a (continuous or discrete) scale from some negative to some positive value, using zero as an average, e.g., for investments in the riskless asset. We could also use a non-negative support for social returns, having some positive number as the average.[11]

With N assets, each asset i can be represented by a tuple of financial and social return, (R_i, S_i). From all assets available, the investor builds the portfolio that maximizes his objective function. Let μ_{R_i} be the expected financial return of asset i, $\sigma^2_{R_i}$ its variance and σ_{R_i,R_j} the covariance between financial returns of assets i and j. R_P denotes the financial return of the portfolio and S_P its social return.

$$R_P = \sum_{i=1}^{N} x_i R_i \qquad \text{and} \qquad S_P = \sum_{i=1}^{N} x_i S_i, \tag{4.1}$$

[9] We claim that there are objective criteria to measure social returns. Thus, the 'warm glow' is only a minor reason to consider social returns. In particular, the social return is then proportional to initial wealth. If the social return comes from the fact that a poor person gets a loan and improves his life (as with a typical investment into microfinance), and the average loan size is a hundred dollar, then one hundred dollars give half the social return of two hundred dollars.

[10] In Dorfleitner, Leidl, and Reeder (2009), we apply the model to real world data and exercise due care regarding social returns. We also comment on the problem when we conclude in Section 4.3.

[11] It does not matter how we scale social returns since they enter the objective function additively, see below.

where the vector $\vec{x} = (x_1, \ldots, x_N)^T$ contains the portfolio weights of the assets with $\sum_i x_i = 1$.

4.2.1 Stochastic social returns

In the introduction, we have argued that, apart from inter-asset dependencies between financial returns, there might be significant inter-asset dependencies of social returns, as well as inter-asset dependencies of financial and social returns. We have also claimed that *intra*-asset dependencies between financial and social returns of an asset might be strong. A case in point for a negative intra-asset relationship is microfinance. The idea of the trade-off is that microfinance borrowers are heterogeneous in quality so that high repayment rates and expected returns are more likely with less poor borrowers. Cull, Demirgüç-Kunt, and Morduch (2007) try to empirically verify the hypothesis of the trade-off and find some evidence in favor of it.[12] In a recent study, Cull, Demirgüç-Kunt, and Morduch (2009, p. 182) write that "[d]ebate also persists on the extent to which trade-offs exist between pursuing profit and reaching the poorest customers. The data here suggests that this trade-off is very real". They also mention the "fear that [...] institutions will sacrifice part of their social missions if subsidies are reduced sharply."

One might object at this point that, even though the trade-off exists, there need not be a stochastic element in it. If MFIs have different strategic orientations - some serving middle-class borrowers, others the very poor - there might only be differences in financial returns and in (deterministic) social returns. However, there is another characteristic feature of the microfinance industry, namely that many MFIs[13] depend on donations.[14] This source of income is highly uncertain in the future and the financial uncertainty might also make social returns risky. An MFI which operates in a market environment and ceases to get funds from donors might be forced to give up on social returns to compensate for the loss of financial returns in order to stay attractive for MFIFs and other investors.[15] In practice, this would be achieved by giving larger loans to less poor people, neglecting the very poor. This has at least

[12] However, having disaggregated data, the authors are able to show that the strength and even the direction of the trade-off depends on institutional design and strategic orientation of the lender.

[13] In principle, the same logic applies to other firms in the realm of donation- or subsidy-receiving industries, as most renewable energy companies, for instance.

[14] From Table 4 in Cull, Demirgüç-Kunt, and Morduch (2009, p. 186), of the 289 MFIs in their sample, the average share of funds that come from donations is 26%.

[15] This chain of arguments assumes the aforementioned trade-off between financial sustainability and outreach.

two theoretical effects which increase financial returns. First, the less poor (and supposedly more educated and productive) clientele might repay more often, and second, giving larger loans reduces transaction costs. The first channel is supported by Morduch (2000, p. 621) who asserts that "[p]roducing and selling goods requires more than just capital. It requires skills, other materials, information, connections, transportation, etc. Since richer households tend to have more of these inputs, marginal returns to capital are often far higher for them than for poorer households".

The concept of 'sin stocks' gives further support to our two central assumptions, namely that social returns matter and that they might have a considerable stochastic component. The literature in this field mainly focuses on whether these allegedly morally doubtful stocks are able to outperform in terms of financial returns. Fabozzi, Ma, and Oliphant (2008) present "empirical evidence that shows sin stocks have outperformed the market on a risk-adjusted basis".[16] First, this result implies that investors do care about social returns. Sin stocks have to offer investors a financial premium to compensate for the 'social harm' done by the investment, vice versa for 'virtue investments'. Second, if a high financial yield of such a company stems from high turnover, their social returns are low when financial returns are high, i.e., there is a negative (intra-asset) covariance, which is an argument in favor of modeling stochastic social returns. This 'success channel' also works the other way round for virtue investing: A solar energy producer has high social returns when output, turnover and profits are high, i.e., a positive intra-asset covariance.

Therefore, we can justify stochastic social returns. But, as mentioned in the introduction, do investors care about variations in social returns? To answer this question, we use the distinction between intrinsic and extrinsic motivation for altruism, as stressed by Sen (1987), for instance. If socially oriented investors are intrinsically motivated, they will care about variations in social returns. After all, they derive utility from the actual results of their investment. Our theoretical model also allows for investors that do care about the expected social return, but not about variations in it (the 'risk-neutral regarding social returns' investors). This type of investor is extrinsically motivated, investing socially in order to tell colleagues, friends

[16] There are also studies rejecting the outperformance hypothesis. Lobe, Roithmeier, and Walkshäusl (2009, p.2) find "results [which] suggest that sin indexes do not offer abnormal returns". An interesting debate about abnormal returns of 'sin stocks' arises in light of the efficient market hypothesis. There should not be an outperformance of sin stocks under rather weak assumptions if markets are efficient. However, this is not the subject of the current paper.

and family about themselves being good guys. In case of corporate investment, it is about telling the public. Their utility stems from telling others, and it arises before (social) returns are actually realized. Therefore, taking the risk of an extreme negative outcome would not decrease their utility much.

Formally, if social returns are stochastic, an investor's objective function depends on two random variables: financial return R_P and social return S_P. In analogy to the financial variables, let μ_{S_i} be the expected social return of asset i, $\sigma^2_{S_i}$ its social variance and σ_{S_i,S_j} the covariance between social returns of assets i and j. If social returns are stochastic, we have cross covariances between financial returns of asset i and social returns of asset j, denoted by σ_{R_i,S_j}. As with individual assets, we use μ_{R_P} as the expected financial return of the portfolio, μ_{S_P} as its expected social return and $\sigma^2_{R_P}$ and $\sigma^2_{S_P}$ as the respective variances of the portfolio. The portfolio covariance is σ_{R_P,S_P}.

Investors' preferences

Consider an investor facing two portfolios. We assume that, first, if both portfolios have identical financial and social means and variances and the same covariance, the investor is indifferent between them, i.e., portfolio choice does not depend on anything else than these five moments. Second, if two portfolios differ only in financial mean, a rational investor prefers the one with the higher mean. Third, the one with the higher social mean is preferred if the portfolios differ only in social mean. Fourth, risk-averse investors choose the portfolio with the lower financial variance if all other things are equal and, fifth, the same is true for two portfolios differing only in social variance. Sixth, since both financial and social returns are a good thing, it makes sense to assume correlation aversion as in Epstein and Tanny (1980). This means that an investor facing two portfolios with the same means and variances prefers the one with the lower (more negative) covariance between financial and social returns. A portfolio with a strong positive covariance is dominated since it pays a low social return exactly when financial returns are low, too. If the covariance is negative, an investor can take comfort in high social returns in times of low financial returns.

Since portfolio choice is an ex ante choice, actual financial and social returns of the portfolio cannot be used to determine the shares of wealth invested in the assets, so that we work with statistical moments. From the literature on the theory of choice under uncertainty, the most common approach is to specify a utility function and then let investors maximize expected utility. This approach is based on the von Neumann-Morgenstern axiomatization, but proves inadequate for many practical applications. As a special case of expected utility maximization, mean-variance analysis gets by with less sophisticated empirical estimates and enjoys great popularity in the finance literature. As is well known, there are ways to guarantee consistency between expected utility and mean-variance in a univariate decision framework. In Appendix 4.4.2, we review these conditions for the univariate case and make a proposal on how to align them in case of a bivariate choice problem, as encountered in our model.

In traditional portfolio theory with only financial returns, Markowitz (1952, p. 82) defines the efficient frontier as the subset of all those mean-variance (E-V) combinations "...with minimum V for given E or more and maximum E for given V or less". Given our assumptions on preferences, we can define the new efficient frontier as follows.

Definition 4.1 *Portfolio A is 'μ_R-μ_S-σ_S-σ_R-$\sigma_{R,S}$-efficient' if there is no portfolio B with*

$$\mu_{R_A} \leq \mu_{R_B},$$

$$\mu_{S_A} \leq \mu_{S_B},$$

$$\sigma_{R_A} \geq \sigma_{R_B},$$

$$\sigma_{S_A} \geq \sigma_{S_B},$$

$$\sigma_{R_A,S_A} \geq \sigma_{R_B,S_B},$$

and at least one inequality strict.

In words, portfolio A is efficient if there is no other portfolio B with at least as high a financial return, at least as high a social return, at most as high a financial risk, at most as high a social risk, and at most as high an *R*-*S* covariance, at least one variable of portfolio B strictly better. In other words, a portfolio is efficient if there is no other, dominant portfolio.[17]

[17]Of course, this definition hinges on the preferences we assume. However, without any assumptions on preferences, an efficient portfolio could be defined as "a portfolio which can be optimal for some kind of preferences", so that every portfolio could be established as an efficient one.

Definition 4.2 *The 'efficient frontier (with stochastic social returns)' is the set of all* μ_R*-*μ_S*-*σ_S*-*σ_R*-*$\sigma_{R,S}$*-efficient portfolios.*

The investor's optimization problem

Given Definition 4.1, the most intuitive way of translating it into an objective function is to simply sum up the five variables and attach weights to them, i.e., to set up an additive preference function[18]

$$\beta_1\mu_{R_P}(\vec{x}) + \beta_2\mu_{S_P}(\vec{x}) - \beta_3\sigma^2_{R_P}(\vec{x}) - \beta_4\sigma^2_{S_P}(\vec{x}) - \beta_5\sigma_{R_P,S_P}(\vec{x}), \tag{4.2}$$

with $\beta_i \geq 0$.[19] Variations in β_i represent changes in preferences. In particular, we have

$\beta_2 = \beta_4 = \beta_5 = 0$: representing a Markowitz investor, social blindness,

$\beta_1 = \beta_3 = \beta_5 = 0$: representing an altruist, social fanatic,

$\beta_1 = \beta_2 = \beta_4 = \beta_5 = 0$, $\beta_3 \neq 0$: representing maximum financial risk aversion.

Consider the following maximization problem

$$\max_{\vec{x}}\{\beta_1\mu_{R_P}(\vec{x}) + \beta_2\mu_{S_P}(\vec{x}) - \beta_3\sigma^2_{R_P}(\vec{x}) - \beta_4\sigma^2_{S_P}(\vec{x}) - \beta_5\sigma_{R_P,S_P}(\vec{x})\}, \tag{4.3}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} x_i = 1.$$

The variables in (4.3) follow from applying standard variance and covariance techniques to the financial and social returns of the portfolio in equation (4.1):

$$\mu_{R_P} = \sum_i x_i\mu_{R_i} \qquad \text{and} \qquad \mu_{S_P} = \sum_i x_i\mu_{S_i}, \tag{4.4}$$

[18] We use 'preference function' instead of 'utility function' since the latter term is usually defined as a function which represents preferences over a set of (deterministic) alternatives, cf. Mas-Colell, Whinston, and Green (1995, p.9). However, some authors talk about 'mean-variance utility', defining a utility function over a set of statistical moments.

[19] The preference function is over-parameterized, but we keep all five parameters for ease of interpretation at this point.

$$\sigma_{R_P}^2 = \vec{x}^T \Sigma_R \vec{x} \qquad \text{and} \qquad \sigma_{S_P}^2 = \vec{x}^T \Sigma_S \vec{x}, \tag{4.5}$$

$$\sigma_{R_P,S_P} = \vec{x}^T \Sigma_{RS} \vec{x}. \tag{4.6}$$

The matrices Σ_R and Σ_S are the covariance matrices

$$\Sigma_R := \begin{pmatrix} \sigma_{R_1,R_1} & \cdots & \sigma_{R_1,R_N} \\ \vdots & \ddots & \vdots \\ \sigma_{R_N,R_1} & \cdots & \sigma_{R_N,R_N} \end{pmatrix}, \qquad \Sigma_S := \begin{pmatrix} \sigma_{S_1,S_1} & \cdots & \sigma_{S_1,S_N} \\ \vdots & \ddots & \vdots \\ \sigma_{S_N,S_1} & \cdots & \sigma_{S_N,S_N} \end{pmatrix}.$$

$$\Sigma_{RS} := \begin{pmatrix} \sigma_{R_1,S_1} & \cdots & \sigma_{R_1,S_N} \\ \vdots & \ddots & \vdots \\ \sigma_{R_N,S_1} & \cdots & \sigma_{R_N,S_N} \end{pmatrix}.$$

The formula for σ_{R_P,S_P} deserves special attention. Apart from the vector $\vec{x}$, it consists of Σ_{RS}, which is a matrix containing two sorts of covariances: intra-asset covariances and inter-asset cross covariances. The diagonal elements σ_{R_i,S_i} of the matrix Σ_{RS} are the intra-asset covariances, which give the dependency between financial and social returns of assets $i = 1, ..., N$. Having called the covariance between financial (or social) returns of different assets the 'normal inter-asset covariance', the off-diagonal elements of Σ_{RS} can be called 'inter-asset cross covariances'. They give the (degree of linear) dependency between social returns of one asset and financial returns of another. In the case of two assets, there are six covariances in total, as shown in Figure 4.1. Note that the normal inter-asset covariances σ_{S_1,S_2} and σ_{R_1,R_2} enter the formulas for the portfolio's financial variance and social variance, respectively.

The solution to our maximization problem in (4.3) is a vector $\vec{x}$, which determines the five preference dimensions (financial and social mean, financial and social variance and covariance). Since $\vec{x}$ is a function of the five preference coefficients β_i, the resulting quintuple is a parameterization of the efficient frontier. We use the Lagrangian

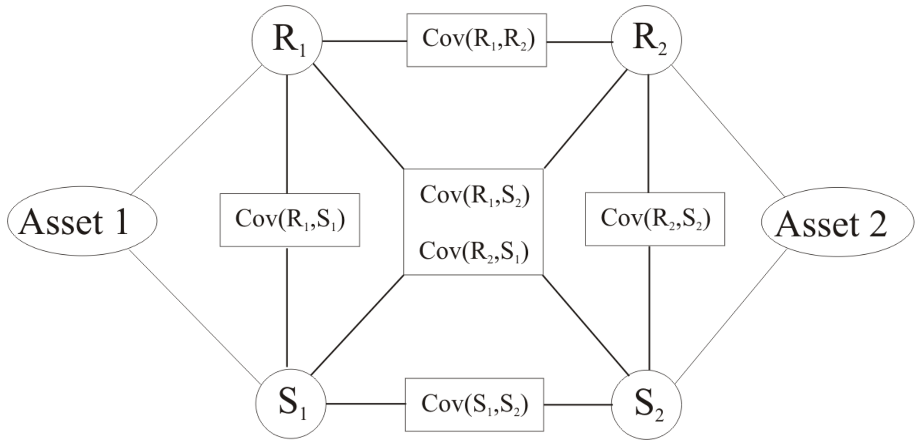


Figure 4.1: All covariances with 2 assets.

$$L(\vec{x},\lambda) \equiv \beta_1\mu_{R_P}(\vec{x})+\beta_2\mu_{S_P}(\vec{x})-\beta_3\sigma^2_{R_P}(\vec{x})-\beta_4\sigma^2_{S_P}(\vec{x})-\beta_5\sigma_{R_P,S_P}(\vec{x})-\lambda\left(\sum_{i=1}^N x_i - 1\right), \quad (4.7)$$

which can be written as

$$\begin{aligned} L(\vec{x},\lambda) &= \beta_1\sum_{i=1}^N x_i\mu_{R_i} + \beta_2\sum_{i=1}^N x_i\mu_{S_i} - \beta_3\sum_{i=1}^N\sum_{j=1}^N x_ix_j\sigma_{R_i,R_j} && (4.8)\\ &- \beta_4\sum_{i=1}^N\sum_{j=1}^N x_ix_j\sigma_{S_i,S_j} - \beta_5\sum_{i=1}^N\sum_{j=1}^N x_ix_j\sigma_{R_i,S_j} - \lambda\left(\sum_{i=1}^N x_i - 1\right). \end{aligned}$$

The first-order condition (FOC) w.r.t. x_i is

$$\frac{\partial L}{\partial x_i} = \beta_1\mu_{R_i}+\beta_2\mu_{S_i}-2\beta_3\sum_{j=1}^N x_j\sigma_{R_i,R_j}-2\beta_4\sum_{j=1}^N x_j\sigma_{S_i,S_j}-\beta_5\left[\sum_{j=1}^N x_j\sigma_{R_i,S_j}+\sum_{j=1}^N x_j\sigma_{R_j,S_i}\right]-\lambda, \quad (4.9)$$

which must be equal to zero for all assets i. The last FOC, $\frac{\partial L}{\partial \lambda}$, is the constraint. Let $\vec{x}^*$ be the vector that contains the optimal asset weights in the portfolio and λ, so that

$$\vec{x}^* := (x_1^*,\ldots,x_2^*,\lambda)^T. \quad (4.10)$$

Moreover, let

$$\vec{y} := (\beta_1\mu_{R_1} + \beta_2\mu_{S_1}, \ldots, \beta_1\mu_{R_N} + \beta_2\mu_{S_N}, 1)^T, \tag{4.11}$$

and

$$M := C_R + C_S + C_{RS} + C_{RS}^T. \tag{4.12}$$

The matrices C_R, C_S and C_{RS} are basically the matrices Σ_R, Σ_S and Σ_{RS} with an additional row and column each, due to the constraint, and the 'preference coefficients'.[20]

$$C_R := \begin{pmatrix} 2\beta_3\sigma_{R_1,R_1} & \ldots & 2\beta_3\sigma_{R_1,R_N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 2\beta_3\sigma_{R_N,R_1} & \ldots & 2\beta_3\sigma_{R_N,R_N} & 1 \\ 1 & \ldots & 1 & 0 \end{pmatrix}, \quad C_S := \begin{pmatrix} 2\beta_4\sigma_{S_1,S_1} & \ldots & 2\beta_4\sigma_{S_1,S_N} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 2\beta_4\sigma_{S_N,S_1} & \ldots & 2\beta_4\sigma_{S_N,S_N} & 0 \\ 0 & \ldots & 0 & 0 \end{pmatrix},$$

$$C_{RS} := \begin{pmatrix} \beta_5\sigma_{R_1,S_1} & \ldots & \beta_5\sigma_{R_1,S_N} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \beta_5\sigma_{R_N,S_1} & \ldots & \beta_5\sigma_{R_N,S_N} & 0 \\ 0 & \ldots & 0 & 0 \end{pmatrix}.$$

Proposition 4.1 *If M is invertible, then*

$$\vec{x}^* = M^{-1}\vec{y} \tag{4.13}$$

is the unique solution to (4.3).

Proof: The matrix M is quadratic, symmetric and of dimension $N+1$. Given that M is invertible, it can be seen that (4.13) is the solution of (4.3). Multiplying equation (4.13) with M from the left and transforming the resulting equation to the underlying system of equations gives the $(N+1)$ FOCs. q.e.d.

The efficient frontier (contingent on our assumptions about preferences) is given in parametric form in equation (4.13). As can be seen from equations (4.11) and (4.12), the solution depends

[20] We could have added the row and column filled with '1' in C_R to any of the other two matrices.

on the five preference parameters $\beta_1, ..., \beta_5$. Let $\mathbb{P} = \mathbb{R}^5_+$ be the Cartesian product of the sets of preference coefficients, where each of the five sets of preference coefficients is the set of positive real numbers. Thus, inserting $\vec{x}^*$ into the five preference dimensions is a mapping from $\mathbb{P}$ to $\mathbb{E}$, the latter being the set of efficient portfolios (with $\mathbb{E} \subset \mathbb{R}^5$). If we were able to plot five-dimensional graphs, we could vary all five parameters from zero to some large number to get a picture of the efficient frontier. In Section 4.2.2, we assume deterministic social returns. This allows us to derive graphs of the efficient frontier.

Liquidity constraints

Usually, investors have some amount of wealth which is free to invest, whereas the majority of their funds is already invested and frequently locked in. In this case, a completely new allocation between assets would have to include transaction costs from premature cancelation. This could be done by subtracting these costs from the financial means. In many other cases, this cost might be prohibitively high. Sometimes, it is not even possible to reallocate, as would be the case for donations (which can also be considered an asset). In these cases, our optimization approach is still applicable, but with modifications concerning the constraints. The ownership of a house worth half a million dollar out of a total wealth of one million pins down, say, x_1 to be one half. The financial and social characteristics of the house must be determined.[21] These values not only pin down x_1 but also impose restrictions on the (financial and social) characteristics of the portfolio. Optimization will then be done over x_i, $i = 2, ..., N$. The change in the objective function can be illustrated with financial mean. Instead of having $\mu_{R_P} = \sum_i x_i \mu_{R_i}$ for all i, we would have to use $\mu_{R_P} = 0.5 \cdot 0.02 + \sum_i x_i \mu_{R_i}, i = 2, ..., N$ if the estimate for the financial mean of the increase in the price of the house is 0.02. The fact that someone has donated money can be incorporated by fixing the respective share (e.g., x_2 to 0.05 for donating $\$10,000$ out of a total wealth of $\$200,000$), using a financial mean of $\mu_{R_2} = -1$, financial risk of zero (the second row and column in Σ_R are zero then) and probably the highest value available of social returns.[22] In the stochastic setup, we should also find variations in social returns of the charitable organization.

[21] In terms of financial returns, this is rather straightforward. Social returns might depend on whether it is a low energy house, how many people live in it per square feet, and so on. Social returns would mainly consist of ecological returns in that case.

[22] Donating money to political parties might have different social returns than giving money to charity.

4.2.2 Deterministic social returns

In this section, we consider a simplified version of the optimization problem (4.3), namely with deterministic social returns. As indicated in the introduction, the resulting problem is similar to the special case of (4.3) which results from setting both β_4 and β_5 equal to zero. This special case would describe an investor that faces stochastic social returns but does not condition its optimal decision on the volatility of social returns. Above, we have called this type of investor 'risk-neutral regarding social returns'.

Preferences with deterministic social returns

Compared to efficiency as defined in Definition 4.1, the search for the efficient frontier becomes easier.

Definition 4.3 *Portfolio A is 'μ-S-σ efficient' if there is no portfolio B with*

$\mu_A \leq \mu_B$,

$S_A \leq S_B$,

$\sigma_A \geq \sigma_B$,

and at least one inequality strict.

Definition 4.4 *The 'efficient frontier (with deterministic social returns)' is the set of all μ-S-σ-efficient portfolios.*

Now, the financial mean and variance of the portfolio are called μ_P and σ_P, respectively. The preference function we use is

$$\beta_1 \mu_P(\vec{x}) + \beta_2 S_P(\vec{x}) - \sigma_P^2(\vec{x}),$$

normalizing the financial risk coefficient, $\beta_3 = 1$. We write S_P instead of μ_{S_P} to stress the fact that social returns are deterministic (whereas the aforementioned special case of the stochastic optimization problem (4.3) with $\beta_4 = \beta_5 = 0$ would use μ_{S_P}).

As in Subsection 4.2.1, we can describe different types of investors by varying the two preference coefficients. In this subsection, risk aversion is unambiguously related to financial risk aversion. Even if there is no canonical or straightforward measure of risk aversion in this setup, we can define the notion of one individual being more risk averse than another in the following way: A tuple $T = (\beta_1, \beta_2)$ represents higher risk aversion than a tuple $T' = (\beta_1', \beta_2')$ if $\beta_1 \leq \beta_1'$ and $\beta_2 \leq \beta_2'$, with strict inequality for one of them. Thus, preferences we are able to model include

$\beta_1 = 0$: financial mean blindness,

$\beta_2 = 0$: social blindness,

$\beta_1 = \beta_2 = 0$: maximum risk aversion,

$\beta_1 \to \infty$ and/or $\beta_2 \to \infty$: risk neutrality.

The set of all possible preference coefficient combinations is now two-dimensional so that the optimal solution $\vec{x}^*$ becomes a mapping from two- into three-dimensional space. This enables us to plot the efficient frontier (see Appendix 4.4.1).

To find the efficient frontier, we consider the simplified optimization problem

$$\max_{\vec{x}} \{\beta_1 \mu_P(\vec{x}) + \beta_2 S_P(\vec{x}) - \sigma_P^2(\vec{x})\}, \tag{4.14}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} x_i = 1.$$

The solution is given in equation (4.13). The matrix M becomes the matrix C_R with $\beta_3 = 1$ since C_S and C_{RS} consist of only zeros in the deterministic case (or, if investors do not care about variations in social returns, $\beta_4 = \beta_5 = 0$ make them consist of only zeros).

The vector $\vec{x}^*$ is a function of the remaining two preference coefficients β_1 and β_2. Inserting the optimal solution $\vec{x}^*$ (without $x_{N+1}^* = \lambda$) into the three portfolio dimensions financial mean, social return and financial risk, the optimal portfolio is a triple depending on two parameters, β_1 and β_2.

The general procedure of optimization also allows for a riskless asset, simply by choosing the characteristics of one of the assets such that it equals the riskless asset, i.e., zero risk and some positive financial return (and a suitable social return). Another way to include a riskless asset is to reformulate the optimization problem. We do this in Appendix 4.4.3.

Tobin-like separation

Definition 4.5 *Preferences of two investors, represented by two tuples* $T = (\beta_1, \beta_2)$ *and* $T' = (\beta'_1, \beta'_2)$ *are said to 'only differ in risk aversion' if, and only if,* $\frac{\beta_1}{\beta_2} = \frac{\beta'_1}{\beta'_2}$ *and* $\beta_j \neq \beta'_j$ *for* $j = 1, 2$.

Let

$$C_R^{-1} := K := \begin{pmatrix} k_{1,1} & \dots & k_{1,N+1} \\ \vdots & \ddots & \vdots \\ k_{N+1,1} & \dots & k_{N+1,N+1} \end{pmatrix}$$

be the inverse of C_R. Thus, we can write the components of the vector $\vec{x}^*$, which maximizes (4.14), as

$$x_i^* = \sum_{j=1}^{N} k_{i,j}(\beta_1 \mu_{R_j} + \beta_2 S_j) + k_{i,N+1}. \tag{4.15}$$

The formula is also valid for $x^*_{N+1} = \lambda$.

This paragraph is concerned with Tobin-like separation properties, i.e., the division of wealth into investment in a riskless asset on the one hand and a risky portfolio on the other. Tobin (1958) considered investors with mean-variance preferences facing a set of risky assets and a riskless asset. Not surprisingly, investors with different attitudes towards risk invest different shares of their total wealth into each of the assets available. However, Tobin was able to show that the portfolio consisting of the risky assets is the same for all investors with μ-σ-preferences, irrespective of their attitudes toward risk. Since this optimal portfolio can be derived as the point of tangency between the riskless asset and the efficient frontier in mean-standard deviation space, Tobin labeled that common optimal portfolio the 'tangent portfolio'.

Proposition 4.2 *There is no single tangent portfolio. The composition of the optimal portfolio of risky assets depends on preferences.*

Proof: Let x_1 be the share of wealth invested in the riskless asset. This determines some values in the matrices C_R and $K = C_R^{-1}$. Clearly, since there is no variation in the return of asset 1, its variance and the covariance with all other assets is zero. Therefore, $\sigma_{R_1,R_j} = \sigma_{R_i,R_1} = 0$ for all $i, j = 1, ..., N$. These changes also affect the inverse of C_R: $k_{i,N+1} = k_{N+1,j} = 0$ for all $i, j \geq 2$ and $k_{1,N+1} = k_{N+1,1} = 1$. Also, we take the riskless asset to have a social return of zero so that $S_1 = 0$.

The portfolio of risky assets consists of the assets $i = 2, ..., N$. The weight of each individual asset in that portfolio can be calculated as

$$x_i^{TP} = \frac{\sum_{j=1}^{N} k_{i,j}(\beta_1 \mu_{R_j} + \beta_2 S_j) + k_{i,N+1}}{\sum_{m=2}^{N} \sum_{j=1}^{N} k_{m,j}(\beta_1 \mu_{R_j} + \beta_2 S_j) + k_{m,N+1}}. \tag{4.16}$$

for $i \geq 2$. It is important to note that normalizing β_3 to unity implies that neither C_R nor K depends on preference coefficients. We can see that variations in β_1 and β_2 affect the weights of the risky assets in the tangent portfolio. q.e.d.

Proposition 4.3 *There is a single optimal portfolio whose composition is identical for investors that only differ in risk aversion.*

Proof: Since $k_{i,N+1} = 0$ for all $i \geq 2$, rearranging equation (4.16), namely by separating sums and dividing numerator and denominator by $\beta_2 (> 0)$, yields:

$$x_i^{TP} = \frac{\frac{\beta_1}{\beta_2} \sum_{j=1}^{N} k_{i,j} \mu_{R_j} + \sum_{j=1}^{N} k_{i,j} S_j}{\frac{\beta_1}{\beta_2} \sum_{m=2}^{N} \sum_{j=1}^{N} k_{m,j} \mu_{R_j} + \sum_{m=2}^{N} \sum_{j=1}^{N} k_{m,j} S_j}. \tag{4.17}$$

Thus, variations in β_1 and β_2 can clearly change the weights in the tangent portfolio. However, a change only in risk aversion according to Definition 4.5 means that we scale both β_1 and β_2 up or down by a constant so that the ratio $\frac{\beta_1}{\beta_2}$ does not change. From equation (4.17), it is only this ratio which influences the weights in the tangent portfolio, not their absolute values. q.e.d.

Proposition 4.3 does not imply that the absolute levels of wealth invested in each of the risky assets do not change. Only the relative weights of the risky assets within the optimal portfolio of risky assets do not change. To see the latter point more clearly, consider the share of total wealth invested in the riskless asset.

$$x_1 = \sum_{j=1}^{N} k_{1,j}(\beta_1 \mu_j + \beta_2 S_j) + 1 = 1 + \beta_1 \sum_{j=1}^{N} k_{1,j}\mu_j + \beta_2 \sum_{j=1}^{N} k_{1,j} S_j, \tag{4.18}$$

since $k_{1,N+1} = 1$. If we scale both β_1 and β_2 up or down such that their ratio $\frac{\beta_1}{\beta_2}$ does not change, x_1 does change. To summarize, even in the deterministic setup, there is no single optimal portfolio, but a set of optimal portfolios instead. Each portfolio in this set represents the optimal portfolio for some type of investor, the type being determined by the preferences in terms of financial returns, risk and social returns. However, we could speak of a conditionally optimal portfolio, conditionality with respect to the relative financial mean and social return preferences: Fixing the ratio $\frac{\beta_1}{\beta_2}$ leads to a single optimal portfolio of risky assets, irrespective of changes in risk aversion. Given such a ratio, asset allocation is a dichotomic problem: What does the optimal portfolio of risky assets look like and how should total wealth be allocated between that portfolio and the riskless asset.

We have seen that Tobin separation breaks down if we include social returns and differences in preferences for social returns into the objective function. In the remainder of this section, we try to give an intuition at the expense of formal rigor. In doing so, we suggest to consider the non-existence of Tobin separation as natural and ask why Tobin separation still obtains if investors only differ in risk aversion.

Consider two types of investors with different preferences. Investors of type 1 do not care about social returns, i.e., $\beta_2 = 0$ while $\beta_1 > 0$. This is the case of standard portfolio theory. By contrast, investors of type 2 have a strong social orientation, such that $\beta_1 = 0$ while $\beta_2 > 0$.[23] Accepting some risk allows both types of investors to gain a premium, both in terms of expected financial returns and social returns. Investor 1 will try to gain a maximal financial premium (at whatever social return) for additional risk, whereas investor 2 will focus on a social premium (at whatever expected financial return) as a compensation for

[23] Some readers might dislike this combination since such investors dislike financial risk, but do not care about the mean of their investment. However, first, we only use this constellation to make our point most clear intuitively, and second, this constellation of preference parameters is analogous to what is assumed to derive the minimum-variance portfolio in standard portfolio theory with only financial returns.

risk. Hence, except for nongeneric cases (e.g., if there is just one risky asset available), their optimal portfolios will differ and Tobin separation breaks down.

We go on to consider three special cases in which Tobin separation obtains nevertheless. First, we consider *different* investors of type 1, i.e., investors with $\beta_2 = 0$ but different $\beta_1 > 0$. Second, we analyze *different* investors of type 2, i.e., with $\beta_1 = 0$ but different $\beta_2 > 0$. Third, we generalize by looking at *different* investors with both $\beta_1 > 0$ and $\beta_2 > 0$. Let these last investors be of type 3.

Type 1 investors are standard Markowitz investors, focusing only on financial returns. It is well known that the optimal portfolio of such an investor maximizes the Sharpe ratio, which is defined as the expected excess return from a portfolio in relation to the portfolio's risk: For any unit of risk borne, the investor tries to get as much additional expected financial return. Let $\mu_{P_{all}}$ denote the expected financial return of an investor's optimal portfolio constructed from all risky assets and the riskless asset. Also, let $\sigma_{P_{all}}$ denote that portfolio's standard deviation. With the riskless rate r, the Sharpe ratio is

$$\frac{\mu_{P_{all}} - r}{\sigma_{P_{all}}}.$$

Both numerator and denominator approach zero for investment in the riskless asset only. Shifting funds toward some portfolio of risky assets changes both numerator and denominator. Since these changes occur linearly for both numerator and denominator, the Sharpe ratio does not change. Thus, different attitudes toward risk (i.e., different $\beta_1 > 0$) make investors of type 1 choose between the portfolio that gives the highest financial premium for a given amount of risk, and the riskless asset. Choosing any other combination of the riskless and/or risky assets leads to a lower Sharpe ratio.[24] This is exactly the reason why Tobin separation holds in a Markowitz world.

Second, consider investors of type 2, the social fanatics. For a given amount of risk, this investor tries to maximize the social premium. Assigning a social return equal to zero to the riskless asset, we can define the 'social Sharpe ratio' as the social excess return per unit of risk,

[24] In fact, the efficient frontier might have linear segments so that there might be infinitely many optimal portfolios. Since this happens in nongeneric cases only, we ignore this subtlety.

$$\frac{S_{P_{all}}}{\sigma_{P_{all}}}.$$

As with the Sharpe ratio, numerator and denominator approach zero for investment in the riskless asset only, but change linearly when funds are shifted between some portfolio and the riskless asset. Assuming that there is a portfolio that maximizes the 'social Sharpe ratio', different social investors of type 2 (with different β_2) will all invest in the same portfolio of risky assets, only the shares of total wealth invested in that portfolio will differ. Thus, we have another special case in which Tobin separation obtains.

Given a social return of the riskless asset equal to zero, we think it is noteworthy that either the financial or the social premium of the optimal portfolio might be negative - but not both at the same time. To see this, consider the extreme situation in which there is only one asset with a positive social return. Then, if this 'social asset' has a financial mean below the riskless rate,[25] an investor of type 2 would choose a portfolio with a high social return at the expense of accepting a negative financial premium and some positive risk. Similar reasoning applies to the financially minded investor 1. However, instead of incurring both negative financial and negative social premia, every investor would prefer to invest in the riskless asset only.

So we have seen that the composition of the optimal portfolios of investors of type 1 and type 2 usually differ. However, we have also seen an intuitive explanation for Tobin separation for different investors of types 1 or 2. Note that both of the above intra-type comparisons can be seen as comparisons of investors that differ in risk aversion only: Since one of the two preference coefficients was assumed to be zero (β_2 for type 1 and β_1 for type 2), changing the absolute value of the other, non-zero preference coefficient does not change the ratio of the two (which is zero or infinity).

Third, let us consider the most interesting type of investors, viz., type 3. We focus on the case of two type-3 investors which only differ in risk aversion. Such investors have different preference coefficients whose ratios are identical though (cf. Definition 4.5). As seen above and independently of the type of investor, additional risk carries two premia, a financial and a social premium. Investors try to maximize the 'social plus financial Sharpe ratio'

[25] This constellation is completely unrealistic and only chosen for ease of exposition. The argument also holds for more realistic situations.

$$SFSR = \frac{\beta_1(\mu_{P_{all}} - r) + \beta_2(S_{P_{all}} - 0)}{\sigma_{P_{all}}},$$

which is a weighted sum of (expected) excess returns per unit of risk incurred. To see why investors maximize this expression, note that investment only in the riskless asset yields a value of the objective function (4.14) of $\beta_1 r$. Increases in risk must be traded-off against increases in the terms which enter the objective function positively. The premium for risk is then $\beta_1\mu_{P_{all}} + \beta_2 S_{P_{all}} - \beta_1 r - \beta_2 0$,[26] or, after rearranging, the numerator of $SFSR$.

All linear combinations of $\mu_{P_{all}}$ and $S_{P_{all}}$, as well as $\sigma_{P_{all}}$ are again linear functions of the share of wealth invested in the riskless asset. Thus, for given β_1 and β_2, shifting funds between the riskless asset and a given portfolio does not change the 'utility premium' per unit of risk. Assuming there is a portfolio which maximizes $SFSR$,[27] differences only in risk aversion will lead to reallocation of funds between the riskless asset and that portfolio. To see this, notice that differences only in risk aversion imply that β_1 and β_2 change such that their ratio does not change. Then, the 'new' $SFSR$ is only a multiple of the 'old' $SFSR$. Therefore, the 'old' optimal portfolio also maximizes the 'new' $SFSR$ and investors which differ only in risk aversion will all invest in that portfolio and the riskless asset, only the respective shares of wealth differ.

Thus, the concepts of Sharpe ratio and 'social Sharpe ratio' are special cases of the $SFSR$. Each of the former two can be considered independently of preference coefficients. Including coefficients would only lead to maximization of some multiple of the original ratio, so that the same optimal portfolio would follow. However, once there are several terms entering the objective function positively and with potentially different coefficients, including preference parameters in order to consider a utility premium becomes indispensable.[28]

4.3 Research outlook and conclusion

What we have done so far can probably be extended to more than just portfolio choice. As mentioned in the introduction to this chapter, we hope to be able to apply the model to asset

[26] Recall that the social return for the riskless asset is assumed to be zero.

[27] Recall that Proposition 4.1 guarantees a unique solution only if the matrix M is invertible.

[28] As a result, interpreting social returns simply as an additional payoff and maximizing $\alpha(\mu_P + S_P) - \gamma\sigma_P^2$ would again lead to Tobin separation. After all, changing α or γ (or both) would only be a special case of investors differing in risk aversion only.

pricing. One way to tractably apply our theoretical results could be to divide investors into two classes: Some purely financially minded, others with some preference for social returns. As a consequence, even if both types of investors differ considerably in terms of risk aversion (only), there would be two optimal portfolios. In the spirit of the CAPM, one could then look for conditions which must hold in order to have an equilibrium in which all investors invest in their respective optimal portfolio of risky assets. Our hope is to be able to derive asset-specific discount factors which can then be applied to future cash-flows in order to predict current market prices. Social returns could then influence asset prices in two possible ways. First, they could be used to derive discount factors which are then applied to financial cash flows (cf. Cochrane, 2005, Ch.1). Second, the standard model of discounting future financial cash-flows could be extended to include both financial and social future cash flows which are then discounted by a financial and social discount factor, respectively. The price of an asset could then be calculated as some mixture between both discounted future cash-flows. However, we have not brought these ideas into a formal framework yet, and it is probably a long way to go from the status quo. In terms of the current model, several interesting extensions are worth further investigations. In terms of economic theory, a desirable robustness check would consist of incorporating measures of risk other than variance. Also, one should impose short-selling restrictions, i.e., add the constraint $x_i \geq 0$ for some i since (in Dorfleitner, Leidl, Reeder, 2009) the theory is applied to investments some of which cannot easily be shorted in reality. This requires the development of an algorithm as in Markowitz (1987). Furthermore, we think that the inclusion of initial wealth of the investor might yield interesting results. The consideration of wealth levels could be such that higher wealth implies less (financial) risk-aversion and more social concern.[29]

Besides further theoretical research, there are interesting applications of the model. To make it more suitable for actual investment decisions, we could transfer modern portfolio techniques like the Black and Litterman (1992) model[30] to the case with social returns.

When putting the model to the data, it is most straightforward to consider MFIFs and

[29] A case in point are the high net worth individuals headed by Bill Gates who started to do financially sustainable investment in the past, but mostly change their behavior going into charity.

[30] The authors recognize the problems associated with the CAPM, namely the need to estimate parameters for all assets available. While asset managers usually have profound knowledge of some assets, they might be completely uninformed about others. Therefore, Black and Litterman develop a technique which allows to specify estimates for only some assets, and then uses CAPM logic to determine estimates for the remaining assets.

other industry-specific funds,[31] where the model can be a guide to help asset managers who have to determine 'the' optimal structure of their products, given a socially-oriented clientele.

Another possible field of interest where investors put emphasis on non-financial objectives is the area of government funds. At this juncture, we might replace the term 'social returns' simply by 'non-financial returns' since governments have a huge range of objectives. Many are concerned with securing stable prices for resources. Most government funds have strategic objectives which include securing future access to resources in foreign countries, or fostering certain industries with special importance.

However, in all areas, the measurement of social returns is still very imperfect and difficult. In some industries, people have proposed scales to measure them. In a very small subset of these, there are also data available. The biggest challenge remains the comparison of social returns between industries. Our model in its most general form with stochastic returns is based on a metric scale of social returns. But comparing social returns of an arms manufacturer and an MFI is very difficult: How bad is it to produce a hundred of guns and how good is it to provide 100 poor children with basic education? And how do we judge all the intermediate companies? Where do we put a software company? How close is the development of an e-learning platform to a microfinance investment? And how far away is the software company programming and distributing ego shooters? These are highly normative questions which cannot be answered using a positive analysis of portfolio choice.

At the same time, we are confident that there will be progress on social metrics. More and more firms become more transparent with regard to social performance. Clearly, firms have an incentive to report on good social performance. The issue then becomes one of proper supervision. Financial intermediaries seeking refinancing from socially oriented investors are putting pressure on their target groups. Reille and Forster (2008, p.15) write that "Triodos [a 'social bank'] is already pushing MFIs to be more transparent about their social impact, lending practices, and environmental policies". The authors also point to the "Global Reporting Initiative" (p.15), which describes itself as "a network-based organization that has pioneered the development of the world's most widely used sustainability reporting framework and is committed to its continuous improvement".[32]

[31] We do this in Dorfleitner, Leidl, and Reeder (2009).

[32] See http://www.globalreporting.org/AboutGRI/WhatIsGRI.

In a way, the burgeoning initiatives on the measurement of social returns reflect our conviction that the paradigm shift of many investors goes on.[33] Thus, the issue of social returns becomes more and more important. In order to properly describe the investment behavior of a growing segment of the market, new economic models are necessary. The model of this chapter can hopefully provide a first step in that direction.

[33]However, we admit that, even if it is possible to measure and compare social returns, it will still be hard to incorporate them into the objective function of an actual investor. One way to derive estimates for the values of β_k is to ask people sophisticated questions which make them reveal their preferences for social returns.

4.4 Appendix

4.4.1 Three-dimensional efficient frontier

Figure 4.2 shows a graph of a three-dimensional efficient frontier.

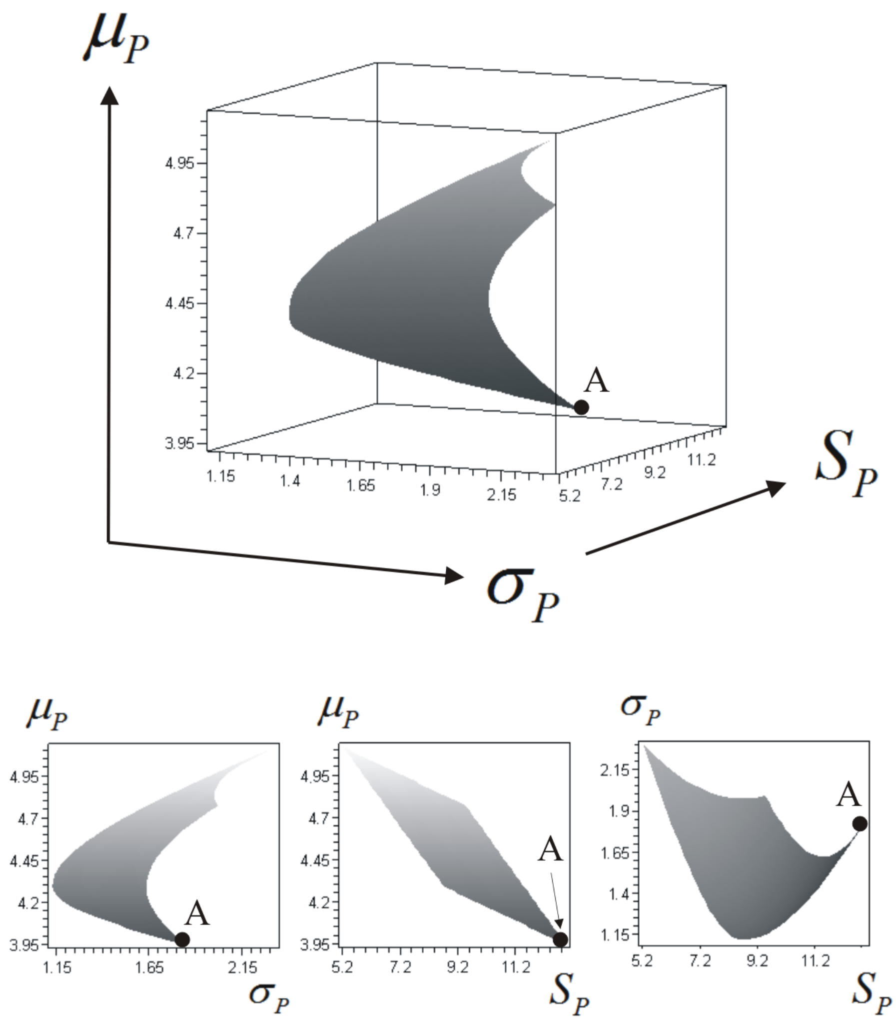


Figure 4.2: Efficient frontier with social returns.

Note that the shape of the efficient frontier is highly sensitive to the set of underlying assets. We used three assets to derive the graph in Figure 4.2: equities (Dow Jones EuroStoxx), bonds (EuroMTS Global Index), and microfinance (responsAbility Global Microfinance Fund). Our estimates for the financial moments can be found in Table 4.1. The equity fund has the highest expected return, viz., 11.01%. The bond fund return is 4.58% and the MFIF's return is only slightly below at 4.05%. In terms of financial risk, the equity fund was most risky with a volatility of 18.52% followed by the bond fund's 3.44% and the MFIF's 1.29%. Both the bond fund and the MFIF exhibit a negative correlation to the equity fund and a positive correlation between each other.

	equity fund	bond fund	MFIF
μ_{R_i}	0.1101	0.0458	0.0405
S_i	0	0	0.01
σ_{R_i}	0.1853	0.0344	0.0129
Correlations			
equity fund	1	-0.2117	-0.3217
bond fund	-0.2117	1	0.1027
MFIF	-0.3217	0.1027	1

Table 4.1: Descriptive statistics.

We vary β_1 and β_2 in the interval $[0, 0.1]$ to get the three-dimensional efficient frontier. The large upper panel in Figure 4.2 shows this surface. The three smaller graphs in the lower panel show the same efficient frontier if we turn it such that one dimension disappears. The first small graph is the three-dimensional efficient frontier in two-dimensional standard mean-variance space. Its outer boundary on the left side resembles the Markowitz frontier. In a Markowitz world, the inner portfolios in that graph would not be efficient since there are portfolios with more return for the same amount of risk and portfolios with less risk for the same return. Even though these 'Markowitz-dominating' portfolios exist here, too, the 'Markowitz-dominated' portfolios can be efficient nonetheless. The reason is that these portfolios have higher social returns than the portfolios on the upper left in the small graph in the lower panel. For instance, consider portfolio A which we have marked in all four graphs. It has a low expected return and a rather high risk. However, it is the portfolio with the highest social return on the surface depicted. Thus, it is the optimal portfolio for an investor with a strong social orientation, i.e., the one characterized by $\beta_1 = 0$ and $\beta_2 = 0.1$.

4.4.2 Expected utility vs. mean-variance

Two excellent surveys about mean-variance justifications can be found in Chapter 4 and 6 in Chavas (2004), and in Chapter 3 in Huang and Litzenberger (1988). For univariate optimization, there are three ways to guarantee consistency of mean-variance and expected utility optimization.

First, consider quadratic utility

$$u(x) = a + bx + cx^2$$

with $b > 0$ and $c < 0$ (risk aversion). Taking expectations yields expected utility

$$EU(x) = a + bE[x] + cE[x^2] = a + bE[x] + c\left(E[x]^2 + Var[x]\right),$$

which is a function of mean and variance of x only. Clearly, quadratic utility has its drawbacks. For consistency with positive marginal utility, x has to be smaller than $\frac{-b}{2c}$. Also, risk aversion requires $c < 0$ which implies increasing absolute risk aversion (IARA), a highly implausible assumption. On the other hand, quadratic utility is a useful proxy for many other utility functions since a second-order Taylor approximation of any differentiable utility function yields a quadratic function.

The second way assumes a specific distribution of returns. Chamberlain (1983) characterizes distribution functions which imply equivalence. Ingersoll (1987, p.96-97) shows that a sufficient condition for equivalence is that returns are multivariate normally distributed.

The third way also assumes a particular distribution. However, it is only the (univariate) portfolio return which needs to be normally distributed. In addition, a specific utility is assumed, namely exponential utility

$$u(x) = -e^{-\gamma x}, \quad \gamma > 0.$$

This function displays the property of constant absolute risk aversion (CARA), which is more realistic than IARA (as with quadratic utility).

Let $\mu = E[X]$ and $\sigma^2 = Var[X]$ be the first two moments of the normally distributed portfolio return x. Since $-\gamma x$ is also normally distributed with $E[-\gamma X] = -\gamma\mu$ and $Var[-\gamma X] = \gamma^2\sigma^2$ the term $e^{-\gamma X}$ is log-normally distributed. Therefore, the expected utility of X is

$$E[-e^{-\gamma X}] = -e^{-\frac{1}{2}(-\gamma^2\sigma^2 + 2\gamma\mu)} = -e^{-\gamma(\mu - \frac{\gamma}{2}\sigma^2)}. \tag{4.19}$$

The last term in the above equation is an increasing function of $\mu - \frac{\gamma}{2}\sigma^2$. Therefore, maximization of $\mu - \frac{\gamma}{2}\sigma^2$ is equivalent to maximizing expected utility. The latter case can also be expanded into the social return dimension when using a biattributive utility function of the form

$$u(x_1, x_2) = -e^{-\gamma_1 x_1} \cdot e^{-\gamma_2 x_2} = -e^{-\gamma_1 x_1 - \gamma_2 x_2}\,.$$

This utility function is of the class $u \circ l$ with u representing a univariate utility function and l a linear form. This class has the comfortable property that the two most important biattributive risk aversion concepts coincide (see Dorfleitner and Krapp (2007)). Assuming a normal distribution of x_1 and x_2, $-\gamma_1 x_1 - \gamma_2 x_2$ is also normally distributed with expectation $-\gamma_1\mu_1 - \gamma_2\mu_2$ and variance $\gamma_1^2\sigma_1^2 + \gamma_2^2\sigma_2^2 + 2\gamma_1\gamma_2\sigma_{1,2}$.[34] Thus, expected utility can be written as

$$Eu(x_1, x_2) = -e^{-\gamma_1\mu_1 - \gamma_2\mu_2 + \frac{1}{2}\gamma_1^2\sigma_1^2 + \frac{1}{2}\gamma_2^2\sigma_2^2 + \gamma_1\gamma_2\sigma_{1,2}}\,.$$

Since $-e^{-\gamma x}$ is an increasing transformation of x, maximization of $\gamma_1\mu_1 + \gamma_2\mu_2 - \frac{1}{2}\gamma_1^2\sigma_1^2 - \frac{1}{2}\gamma_2^2\sigma_2^2 - \gamma_1\gamma_2\sigma_{1,2}$ is identical to maximization of expected utility.

4.4.3 Tobin separation: an alternative approach

We again make the simplifying assumption of deterministic social returns.

Instead of interpreting one of the assets in our above optimization problem as the riskless one, explicit modeling is even easier. So let x_0 be the share of wealth invested in the riskless asset, with $x_0 = 1 - \sum_{i=1}^{N} x_i$ being the residual share. As in the above section, apart from having a variance of zero per se, we assume that the riskless asset has no social return, i.e., $S_0 = 0$.

The portfolio financial mean μ_P becomes

$$\mu_P = r + \sum_{i=1}^{N} x_i(\mu_i - r)\,.$$

The optimization problem is

[34] μ_j and σ_j^2 are the expectation and variance of x_j, respectively. $\sigma_{1,2}$ is the covariance between x_1 and x_2.

$$\max_{x_1,\ldots,x_N} \beta_1 r + \beta_1 \sum_{i=1}^{N} x_i(\mu_i - r) + \beta_2 \sum_{i=1}^{N} x_i S_i - \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{R_i,R_j}\,.$$

We do not need a constraint on the sum of the asset weights in the portfolio. By construction, optimal weights $x_0, ..., x_N$ add up to one. Again differentiating w.r.t. the weights $\vec{x}$, optimal weights are

$$\vec{x}^* = C^{-1} \times \vec{z}\,,$$

where $\vec{x}^*$ is now of dimension N since there is no constraint. We use

$$\vec{z} := (\beta_1(\mu_1 - r) + \beta_2 S_1, \ldots, \beta_1(\mu_N - r) + \beta_2 S_N)^T$$

(of dimension N, too) and the $(N \times N)$ matrix

$$C := \begin{pmatrix} 2C_{11} & \ldots & 2C_{1N} \\ \vdots & \ddots & \vdots \\ 2C_{N1} & \ldots & 2C_{NN} \end{pmatrix} = 2\Sigma_R\,.$$

Let

$$C^{-1} = K := \begin{pmatrix} k_{11} & \ldots & k_{1N} \\ \vdots & \ddots & \vdots \\ k_{N1} & \ldots & k_{NN} \end{pmatrix},$$

so that the weights x_i for $i = 1, ..., N$ follow

$$x_i = \beta_1 \sum_{j=1}^{N} k_{ij}(\mu_j - r) + \beta_2 \sum_{j=1}^{N} k_{ij} S_j\,.$$

Excluding the case of $\beta_2 = 0$ (the Markowitz case), we get

$$x_i^{TP} = \frac{\frac{\beta_1}{\beta_2}\sum_{j=1}^{N} k_{ij}(\mu_j - r) + \sum_{j=1}^{N} k_{ij} S_j}{\frac{\beta_1}{\beta_2}\sum_{m,j=1}^{N} k_{mj}(\mu_j - r) + \sum_{m,j=1}^{N} k_{mj} S_j}\,.$$

The share of asset i in the optimal portfolio of risky assets only depends on the ratio $\frac{\beta_1}{\beta_2}$.

Chapter 5

Conclusion

This dissertation is the result of three years of research. Prior to and during that time, we have observed three facts which seemed to be underrepresented in economic theory. First, markets in general and financial markets in particular seem to be characterized by the existence of immense aggregate risks. Second, formal credit markets in developing countries grew, and many MFIs became able to detach themselves from governments and donor agencies, seeking market refinancing instead. Third, and certainly not independent from the previous fact, the emergence of 'social funds' suggested that investors increasingly take into account non-financial objectives when they decide about how to invest. Our aim was to develop a microeconomic foundation which more closely represents these stylized facts. In order to do so, we resorted to three well-established formal frameworks.

In Chapter 2, we extended the seminal Stiglitz and Weiss (1981) model by introducing dependent project revenues of risky firms. We have pointed out the conditions under which households react in one way or another to the capital risk implied by the dependency of firms' revenues. We have seen that the households' risk attitude is crucial to determine the credit market equilibrium. Establishing a notion of social optimum in the presence of aggregate risk, we were able to show that both overinvestment and underinvestment are possible in the market equilibrium. In a comparison to the SW model with independent project revenues, we have seen that project dependency can aggravate adverse selection in that it reduces the amount of active safe firms. In light of the forthcoming publication of Arnold and Riley (2009), we found another interesting relationship. The impossibility of a globally hump-shaped expected return function does not imply that credit rationing is impossible in equilibrium, i.e., a hump-shaped return function is not a necessary condition for credit rationing.

Chapter 3 is based on one of the most often cited models in the realm of microfinance, viz., the strategic default model of Besley and Coate (1995). In a first step, we took their repayment game as it is, i.e., with non-cooperative behavior of borrowers, and analyzed its performance in a credit market equilibrium. We have seen that the equilibrium might consist of both individual and group lending contracts at the same time if penalties are completely pecuniary. Considering the more natural assumption of non-pecuniary penalties, we found a counterintuitive result: The equilibrium might be characterized by individual lending contracts even though group lending contracts would yield the higher repayment rate and could be offered at lower loan rates. As an explanation to this phenomenon, we identified inefficiently high penalties through banks, which can be attributed to borrowers' non-cooperative behavior. Assuming that borrowers cooperate in the repayment game, this odd result disappeared. More than that, we were able to show that group lending is the unique mode of finance when penalties are non-pecuniary and borrowers cooperate. We also considered the effects of social sanctions. Even though sanctions increase the repayment rate of group lending both compared to individual lending and to group lending without sanctions, they can have critical implications. In order to avoid stigmatization through social sanctions, borrowers might sell their belongings to repay a group loan, even though the bank is unable to force borrowers to do so. In that sense, social sanctions can undermine limited liability. The acceptance of a group lending contract might then drive borrowers into deeper poverty. If borrowers are risk-averse, they might refuse such contracts with consequences for the market equilibrium. This can also happen if a part of the penalty is pecuniary. By contrast, if borrowers are risk-neutral and if penalties are non-pecuniary, we have shown that any loan rate that banks would possibly offer in equilibrium implies the use of group lending contracts and the participation of borrowers. As a general result, our model suggests that the kind of market failures known from the literature on asymmetric information are also present in the microcredit market characterized by enforcement problems, irrespective of both the degree of cooperation between borrowers and the existence of social sanctions. Cooperation and social sanctions ameliorate but do not eliminate these market failures.

In Chapter 4, we extended Markowitz (1952) to include a social dimension. We argued in favor of including social returns into an investor's objective function and solved the resulting optimization problem, both for stochastic and deterministic social returns. We defined the new efficient frontier and derived two interesting properties: In general, optimal portfolios of different investors are not the same if the investors' preferences are represented by the extended mean-variance function. However, if investors only differ in risk aversion, we have seen that there is a single optimal portfolio for all such investors. The optimal solution as a function of the investors' preferences allowed us to determine the optimal structure of different investors' portfolios. Moreover, asset managers of social funds can use the solution of the model with stochastic social returns to determine an optimal allocation of wealth. For given investor preferences, the solution shows how individual assets should be combined in order to maximize utility arising as a combination of social and financial returns, taking into account correlations between social and financial returns of the assets.

It was interesting to see how our understanding of the phenomena considered has evolved over time. Some of our results confirm what would have been expected intuitively. When we started thinking about aggregate risks in the SW model, we were convinced that credit rationing is a likely outcome. However, the number of channels and mechanisms involved grew steadily. We were fascinated by the diversity of results following a small change in the utility function or in the assumptions regarding the social optimum. We hope to have provided an exposition which helps the reader understand the subtleties involved when thinking about the effects of aggregate risks in a credit market equilibrium. We expect to see further research on this topic in more advanced models.

By contrast, while thinking about the incorporation of social returns in a model of portfolio choice, we had no expectations as to where we would go. We have developed a simple framework which we can hopefully extend to answer questions related to asset pricing. This becomes particularly important if microfinance continues to grow at current rates. In order to ensure that the global fight against poverty is advancing as quickly and efficiently as possible, further research on microcredit markets is highly desirable.

Bibliography

Ahlin, Christian, and Neville Jiang, 2008, Can micro-credit bring development?, *Journal of Development Economics* 86, 1–21.

Ahlin, Christian, and Robert M. Townsend, 2002, Using Repayment Data to Test Across Models of Joint Liability Lending, *working paper*.

——— , 2007, Using Repayment Data to Test Across Models of Joint Liability Lending, *Economic Journal* 117(517), F11–F51.

Akerlof, George A., 1970, The Market for ”Lemons”: Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics* 84, 488–500.

Armendáriz de Aghion, Beatriz, and Christian Gollier, 2000, Peer Group Formation in an Adverse Selection Model, *Economic Journal* 110(465), 632–43.

Armendáriz de Aghion, Beatriz, and Jonathan Morduch, 2005, *The Economics of Microfinance* (The MIT Press).

Arnold, Lutz G., 2007, A Game-Theoretic Foundation for the Stiglitz-Weiss Model, *Manuscript, University of Regensburg.*

——— , Johannes J. Reeder, and Susanne Steger, 2009, Microfinance and markets: New results for the Besley-Coate group lending model, *University of Regensburg Working Papers in Business, Economics and Management Information Systems.*

Arnold, Lutz G., Johannes J. Reeder, and Stefanie Trepl, 2010, Aggregate Risk in the Stiglitz-Weiss Model, *working paper.*

Arnold, Lutz G., and John G. Riley, 2009, On the Possibility of Credit Rationing in the Stiglitz-Weiss Model, *American Economic Review* 99(5), 2035–44.

Attanasio, Orazio P., and Guglielmo Weber, 1989, Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption, *Economic Journal* 99, 59–73.

Baltensperger, Ernst, 1978, Credit Rationing: Issues and Questions, *Journal of Money, Credit, and Banking* 10(2), 170–83.

Banerjee, Abhijit V., Timothy Besley, and Timothy W. Guinnane, 1994, Thy Neighbor's Keeper: The Design of a Credit Cooperative with a Theory and a Test, *Quarterly Journal of Economics* 109(2), 491–515.

Banerjee, Abhijit V., and Esther Duflo, 2007, The Economic Lives of the Poor, *Journal of Economic Perspectives* 21(1), 141–68.

Basu, Parantap, and Satyajit Ghosh, 1993, Optimal Saving under General Changes in Uncertainty: A Nonexpected Utility Maximizing Approach, *Southern Economic Journal* 60, 119–27.

Berger, Allen N., and Gregory F. Udell, 1992, Some Evidence on the Empirical Significance of Credit Rationing, *Journal of Political Economy* 100(5), 1047–77.

Bernhardt, Dan, 2000, Credit Rationing?, *American Economic Review* 90(1), 235–39.

Besanko, David, and Anjan V. Thakor, 1987b, Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Credit Markets, *International Economic Review* 28(3), 671–89.

Besley, Timothy J., and Stephen Coate, 1995, Group lending, repayment incentives and social collateral, *Journal of Development Economics* 46, 1–18.

Bester, Helmut, 1985, Screening vs. Rationing in Credit Markets with Imperfect Information, *American Economic Review* 75(4), 850–55.

———, 1994, The Role of Collateral in a Model of Debt Renegotiation, *Journal of Money, Credit and Banking* 26(1), 72–86.

——— , and Martin F. Hellwig, 1987, Moral Hazard and Equilibrium Credit Rationing: An Overview of the Issues, in Günter Bamberg, and Klaus Spremann, ed.: *Agency Theory, Information, and Incentives,* pp. 135–66. Springer.

Bhattacharya, Sudipto, Arnoud Boot, and Anjan V. Thakor, 2004, *Credit Markets, Intermediation, and the Macroeconomy* (Oxford University Press).

Bhattacharya, Sudipto, and Anjan V. Thakor, 1993, Contemporary Banking Theory, *Journal of Financial Intermediation* 3, 2–50.

Bhole, Barat, and Sean Ogden, 2009, Group lending and individual lending with strategic default, *Journal of Development Economics.*

Black, Fischer, and Robert Litterman, 1992, Global Portfolio Optimization, *Financial Analysts Journal* 48(5), 28–43.

Bond, Philip, and Ashok Rai, 2002, Collateral Substitutes in Microfinance, *mimeo, Harvard university.*

Breeden, Douglas T., 1979, An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities, *Journal of Financial Economics* 7(3), 227–318.

Carr, James H., and Zhong Yi Tong, 2002, *Replicating Microfinance in the United States* (John Hopkins University Press).

Cassar, Alessandra, Luke Crowley, and Bruce Wydick, 2007, The Effect of Social Capital on Group Loan Repayment: Evidence from Field Experiments, *Economic Journal* 117(517), F85–F106.

Chamberlain, Gary, 1983, A Characterization of the Distributions That Imply Mean-Variance Utility Functions, *Journal of Economic Theory* 29(1), 185–201.

Chan, Yuk-Shee, and George Kanatas, 1985, Asymmetric Valuations and the Role of Collateral in Loan Agreements, *Journal of Money, Credit and Banking* 17(1), 84–95.

Chavas, Jean-Paul, 2004, *Risk Analysis in Theory and Practice* (Elsevier Academic Press).

Clemenz, Gerhard, 1986, *Credit Markets with Asymmetric Information* (Springer).

Cochrane, John H., 2005, *Asset Pricing* (Princeton University Press).

Conning, Jonathan, 2005, Monitoring by Delegates or by Peers? Joint Liability Loans under Moral Hazard, Hunter College Department of Economics Working Papers 407.

Cull, Robert, Asli Demirgüç-Kunt, and Jonathan Morduch, 2007, Financial Performance and Outreach: A Global Analysis of Leading Microbanks, *Economic Journal* 117, F107–F133.

——— , 2009, Microfinance Meets the Market, *Journal of Economic Perspectives* 23(1), 167–92.

D'Antonio, Louis, and Tommi Johnsen, 1997, Expanding Socially Screened Portfolios: An Attribution Analysis of Bond Performance, *Journal of Investing* 6, 79–87.

——— , and Bruce Hutton, 2000, Socially Responsible Investing and Asset Allocation, *Journal of Investing* 9, 65–73.

De Meza, David, and David C. Webb, 1987, Too Much Investment: A Problem of Asymmetric Information, *Quarterly Journal of Economics* 102(2), 281–92.

——— , 1992, Efficient credit rationing, *European Economic Review* 36, 1277–90.

——— , 2006, Credit Rationing: Something's Gotta Give, *Economica* 73, 563–78.

Diamond, Douglas W., 1984, Financial Intermediation and Delegated Monitoring, *Review of Economic Studies* 51(3), 393–414.

——— , 1989, Reputation Acquisition in Debt Markets, *Journal of Political Economy* 97(4), 828–62.

Dieckmann, Raimar, 2007, Microfinance: An emerging investment opportunity, Current issues Deutsche Bank Research.

Dorfleitner, Gregor, and Michael Krapp, 2007, On multiattributive risk aversion: some clarifying results, *Review of Managerial Science* 1, 47–63.

Dorfleitner, Gregor, Michaela Leidl, and Johannes J. Reeder, 2009, Theory of Social Returns in Portfolio Choice with Application to Microfinance, unpublished.

Dunn, Brian, 2006, Modern Portfolio Theory - with a Twist: The New Efficient Frontier, *http://americansforcommunitydevelopment.org/supportingdownloads/AcquillianPresentation07.pdf*.

Dupré, Denis, Isabelle Girerd-Potin, and Raghid Kassoua, 2004, Adding an ethical dimension to portfolio management, *Finance India* 18, 625–42.

Eichberger, Jürgen, and Ian R. Harper, 1997, *Financial Economics* (Oxford University Press).

Elsas, Ralf, and Jan P. Krahnen, 2004, Universal Banks and Relationships with Firms, in Jan P. Krahnen, and Reinhard H. Schmidt, ed.: *The German Financial System*. Oxford University Press.

Epstein, Larry G., and Stephen M. Tanny, 1980, Increasing Generalized Correlation: a Definition and some Economic Consequences, *Canadian Journal of Economics* 13, 16–34.

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework, *Econometrica* 57(4), 937–69.

——— , 1991, Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis, *Journal of Political Economy* 99(2), 263–86.

Fabozzi, Frank J., K.C. Ma, and Becky J. Oliphant, 2008, Sin Stock Returns, *Journal of Portfolio Management* 35(1), 82–94.

Fama, Eugene F., and Kenneth R. French, 1995, Size and Book-to-Market Factors in Earnings and Returns, *Journal of Finance* 50(1), 131–55.

Ferreira, Miguel A., and Paulo M. Gama, 2005, Have World, Country, and Industry Risks Changed over Time? An Investigation of the Volatility of Developed Stock Markets, *Journal of Financial and Quantitative Analysis* 40(1), 195–222.

Freixas, Xavier, and Jean-Charles Rochet, 2008, *Microeconomics of Banking* (The MIT Press).

Gale, Douglas, and Martin F. Hellwig, 1985, Incentive-Compatible Debt Contracts: The One-Period Problem, *Review of Economic Studies* 52(4), 647–63.

Ghatak, Maitreesh, 1999, Group lending, local information, and peer selection, *Journal of Development Economics* 60(1), 27–50.

——— , 2000, Screening by the Company You Keep: Joint Liability Lending and the Peer Selection Effect, *Economic Journal* 110(465), 601–31.

——— , and Timothy W. Guinnane, 1999, The economics of lending with joint liability: theory and practice, *Journal of Development Economics* 60, 195–228.

Giné, Xavier, and Dean S. Karlan, 2009, Group versus Individual Liability: Long Term Evidence from Philippine Microcredit Lending Groups, Working papers Economic Growth Center, Yale University.

Gomez, Rafael, and Eric Santor, 2003, Do Peer Group Members Outperform Individual Borrowers? A Test of Peer Group Lending Using Canadian Micro-Credit Data, Working Papers 03-33, Bank of Canada.

Gonzalez, Adrian, 2007, Efficiency Drivers of Microfinance Institutions (MFIs): The Case of Operating Costs, *MicroBanking Bulletin* 15, 37–42.

Goodhart, Charles A.E., 2004, *Financial Development and Economic Growth: Explaining the Links* (Palgrave Macmillan).

Greenwald, Bruce C., and Joseph E. Stiglitz, 1993, Financial Market Imperfections and Business Cycles, *Quarterly Journal of Economics* 108(1), 77–114.

Grootveld, Henk, and Winfried Hallerbach, 1999, Variance vs downside risk: Is there really that much difference?, *European Journal of Operational Research* 114(2), 304–19.

Hall, Robert E., 1987, Consumption, *NBER Working Paper Series* No. 2265.

Hart, Oliver D., and Bengt Holmström, 1987, The theory of contracts, in Truman F. Bewley, ed.: *Advances in Economic Theory Fifth World Congress* pp. 71–155. Cambridge University Press.

Hellmann, Thomas, and Joseph E. Stiglitz, 2000, Credit and equity rationing in markets with adverse selection, *European Economic Review* 44, 281–304.

Hillier, Brian, and Muradali V. Ibrahimo, 1993, Asymmetric Information and Models of Credit Rationing, *Bulletin of Economic Research* 45(4), 271–304.

Hoepner, Andreas G.F., 2007, A categorisation of the responsible investment literature, Discussion paper, SSRN.

Hogg, Michael A., Deborah J. Terry, and Katherine M. White, 1995, A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory, *Social Psychology Quarterly* 58(4), 255–69.

Holmström, Bengt, and Paul Milgrom, 1990, Regulating Trade Among Agents, *Journal of Institutional and Theoretical Economics* 146, 85–105.

Huang, Chi-fu, and Robert H. Litzenberger, 1988, *Foundations for Financial Economics* (North-Holland).

Hubbard, Douglas W., 2007, *How to measure anything: Finding the value of 'intangibles' in business* (John Wiley).

Ingersoll, Jonathan E., 1987, *Theory of Financial Decision Making* (Rowman & Littlefield).

Jaffee, Dwight M., 1971, *Credit Rationing and the Commercial Loan Market: An Econometric Study of the Structure of the Commercial Loan Market* (John Wiley).

——— , and Thomas Russell, 1976, Imperfect Information, Uncertainty, and Credit Rationing, *Quarterly Journal of Economics* 90, 651–66.

Keeton, William R., 1979, *Equilibrium Credit Rationing* (Garland Publishing, Inc.).

Knight, Frank H., 1967, *Risk, Uncertainty and Profit* (Kelley).

Kreps, David M., 1997, Intrinsic Motivation and Extrinsic Incentives, *American Economic Review* 87(2), 359–364.

Laffont, Jean-Jacques, 2003, Collusion and group lending with adverse selection, *Journal of Development Economics* 70(2), 329–48.

——— , and David Martimort, 2002, *The Theory of Incentives: The Principal-Agent Model* (Princeton University Press).

Laffont, Jean-Jacques, and Tchétché N'Guessan, 2000, Group Lending with Adverse Selection, *European Economic Review* 44(4-6), 773–84.

Laffont, Jean-Jacques, and Patrick Rey, 2003, Moral Hazard, Collusion and Group Lending, IDEI Working Papers, Institut d'Économie Industrielle (IDEI), Toulouse.

Lintner, John, 1965, The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Assets, *Review of Economics and Statistics* 47(1), 13–37.

Lobe, Sebastian, Stefan Roithmeier, and Christian Walkshäusl, 2009, Vice vs. Virtue investing, Discussion paper, SSRN.

Lucas, Robert E., Jr., 1978, Asset Prices in an Exchange Economy, *Econometrica* 46(6), 1429–45.

Luenberger, David G., 1998, *Investment Science* (Oxford University Press).

Mankiw, N. Gregory, 1986, The Allocation of Credit and Financial Collapse, *Quarterly Journal of Economics* 101(3), 455–70.

Markowitz, Harry M., 1952, Portfolio Selection, *Journal of Finance* 7, 77–91.

——— , 1987, *Mean-Variance Analysis in Portfolio Choice and Capital Markets* (Blackwell).

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green, 1995, *Microeconomic Theory* (Oxford University Press).

Merton, Robert C., 1971, Optimum Consumption and Portfolio Rules in a Continuous-Time Model, *Journal of Economic Theory* 3(4), 373–413.

——— , 1973, An Intertemporal Capital Asset Pricing Model, *Econometrica* 41(5), 867–87.

Milde, Hellmuth, and John G. Riley, 1988, Signaling in Credit Markets, *Quarterly Journal of Economics* 103(1), 101–29.

Morduch, Jonathan, 1999, The Promise of Microfinance, *Journal of Economic Literature* 37(4), 1569–1614.

——— , 2000, The Microfinance Schism, *World Development* 28(4), 617–29.

——— , 2006, Microinsurance: The Next Revolution?, in Abhijit V. Banerjee, Roland Bénabou, and Dilip Mookherjee, ed.: *Understanding poverty,* pp. 337–56. Oxford University Press.

Mossin, Jan, 1966, Equilibrium in a Capital Asset Market, *Journal of the Econometric Society* 34(4), 768–783.

Petersen, Mitchell A., and Raghuram G. Rajan, 1995, The Effect of Credit Market Competition on Lending Relationships, *Quarterly Journal of Economics* 110(2), 407–43.

——— , 1997, Trade Credit: Theories and Evidence, *Review of Financial Studies* 10(3), 661–92.

Putnam, Robert D., 1993, The Prosperous Community - Social Capital and Public Life, *American Prospect* 13.

Rai, Ashok S., and Tomas Sjöström, 2004, Is Grameen Lending Efficient? Repayment Incentives and Insurance in Village Economies, *Review of Economic Studies* 71(1), 217–34.

Reeder, Johannes J., and Susanne Steger, 2008, Microcredit: Besley and Coate's repayment game in a credit market equilibrium with pecuniary penalties, unpublished.

Reeder, Johannes J., and Stefanie Trepl, 2009, Dependent Revenues, Capital Risk and Credit Rationing, *BGPE discussion paper series* No. 78.

Reichlin, Pietro, 2004, Credit Markets, Intermediation, and the Macroeconomy: A Discussion, in Sudipto Bhattacharya, Arnoud Boot, and Anjan V. Thakor, ed.: *Credit Markets, Intermediation, and the Macroeconomy* pp. 856–91. Oxford University Press.

Reille, Xavier, and Sarah Forster, 2008, Foreign Capital Investment in Microfinance, Focus Note 44, CGAP.

Reille, Xavier, Jasmina Glisovic-Mezieres, Yanniz Berthouzoz, and Damian Milverton, 2009, MIV Performance and Prospects: Highlight from the CGAP 2009 MIV Benchmark Survey, Brief, CGAP.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang, 2008, Socially responsible investments: Institutional aspects, performance, and investor behavior, *Journal of Banking and Finance* 32(9), 1723–42.

Rhyne, Elisabeth, 2009, *Microfinance for bankers and investors: unterstanding the opportunities and challenges of the market at the bottom end of the pyramid* (McGraw Hill).

Riley, John G., 1987, Credit Rationing: A Further Remark, *American Economic Review* 77(1), 224–27.

Rosa, Robert V., 1951, Interest Rates and the Central Bank, in *Money, Trade, and Economic Growth: In Honor of John Henry Williams* pp. 270–95. The Macmillan Company.

Ross, Stephen A., 1976, The Arbitrage Theory of Capital Asset Prices, *Journal of Economic Theory* 13(3), 341–60.

———, Randolph W. Westerfield, and Jeffrey Jaffe, 2005, *Corporate Finance* (McGraw-Hill).

Rothschild, Michael, and Joseph E. Stiglitz, 1971, Increasing Risk II: Its Economic Consequences, *Journal of Economic Theory* 3(1), 66–84.

Scott, Ira O. Jr., 1957, The Availability Doctrine: Theoretical Underpinnings, *Review of Economic Studies* 25(1), 41–48.

Selden, Larry, 1978, A New Representation of Preferences over "Certain × Uncertain" Consumption Pairs: The "Ordinal Certainty Equivalent" Hypothesis, *Econometrica* 46(5), 1045–60.

———, 1979, An OCE Analysis of the Effect of Uncertainty on Saving under Risk Preference Independence, *Review of Economic Studies* 46(1), 73–82.

Sen, Amartya, 1987, *On Ethics and Economics* (Basil Blackwell Ltd).

Sharpe, William F., 1964, Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk, *Journal of Finance* 19(3), 425–42.

Slovin, Myron B., and Marie E. Sushka, 1983, A Model of the Commercial Loan Rate, *Journal of Finance* 38(5), 1583–96.

Stahl, Dale O., 1988, Bertrand Competition for Inputs and Walrasian Outcomes, *American Economic Review* 78(1), 189–201.

Stiglitz, Joseph E., 1977, Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market, *Review of Economic Studies* 44(3), 407–30.

———, 1990, Peer monitoring and credit markets, *World Bank Economic Review* 4, 351–66.

———, 2000, The Contributions Of The Economics Of Information To Twentieth Century Economics, *Quarterly Journal of Economics* 115(4), 1441–1478.

———, 2002, Information and the Change in the Paradigm in Economics, *American Economic Review* 92(3), 460–501.

———, and Andrew Weiss, 1981, Credit Rationing in Markets with Imperfect Information, *American Economic Review* 71(3), 393–410.

———, 1983, Incentive Effects of Terminations: Applications to the Credit and Labour Markets, *American Economic Review* 73(5), 912–27.

Tobin, James, 1958, Liquidity Preference as Behavior Towards Risk, *Review of Economic Studies* 67, 65–86.

Townsend, Robert M., 1979, Optimal Contracts and Competitive Markets with Costly State Verification, *Journal of Economic Theory* 21(2), 265–93.

———, 2003, Microcredit and Mechanism Design, *Journal of the European Economic Association* 1, 468–77.

Van Tassel, Eric, 1999, Group lending under asymmetric information, *Journal of Development Economics* 60(1), 3–25.

Varian, Hal R., 1990, Monitoring Agents With Other Agents, *Journal of Institutional and Theoretical Economics* 146(1), 153–74.

von Neumann, John, and Oskar Morgenstern, 1944, *Theory of Games and Economic Behavior* (Princeton University Press).

Williamson, Stephen D., 1986, Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing, *Journal of Monetary Economics* 18(2), 159–79.

——— , 1987, Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing, *Quarterly Journal of Economics* 102(1), 135–45.

Worchel, Stephen, Hank Rothgerber, Eric A. Day, D. Hart, and John Butemeyer, 1998, Social identity and individual productivity within groups, *British journal of social psychology* 37(4), 389–413.

Yanelle, Marie-Odile, 1989, The Strategic Analysis of Intermediation, *European Economic Review* 33, 294–301.

——— , 1997, Banking Competition and Market Efficiency, *Review of Economic Studies* 64, 215–39.

Yunus, Muhammad, 2003, *Banker to the Poor* (Perseus Books Group).

——— , 2007, *Creating a World Without Poverty: Social Business and the Future of Capitalism* (Perseus Books Group).