Three Essays in the Theory of Contracts and Organizations

by

Fausto Panunzi

Laurea, Universita' Bocconi, Milan (1988)

Submitted to the Department of Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1995

© Fausto Panunzi, MCMXCV

The author hereby grants to MIT permission to reproduce and to distribute copies of this thesis document in whole or in part.

Signature of Author .................................................................

Department of Economics

March 16, 1995

Certified by ..............................................................

Oliver Hart
Professor of Economics, Harvard University
Thesis Supervisor

Certified by ..........

Jean Tirole
Professor of Economics
Thesis Supervisor

Accepted by .................................................................

Richard Eckaus
Chairman, Departmental Committee on Graduate Students
Three Essays in the Theory of Contracts and Organizations

by

Fausto Panunzi

Submitted to the Department of Economics
on March 16, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The first chapter analyzes how the ownership structure of a firm can act as a commitment device to delegate a certain degree of authority from the shareholders to the management. While a dispersed ownership grants the management much discretionary power, a more concentrated one results in closer monitoring and hence less initiative by the manager. To the extent that managerial initiative is beneficial for the shareholders there is a trade-off between the gains from monitoring and those from managerial activism. This trade-off is analyzed successively in the case of aligned and conflicting shareholders’ interests. In both cases the optimal size of the large shareholder’s block balances the conflicting effects of monitoring. These results are applied to the analysis of fiduciary duty and signal manipulation providing new insights on traded subsidiaries and predation.

In the second chapter we apply the theory developed in the first chapter to the study of the design of contracts in a borrower-lender relationship. More precisely, we investigate how differences in the lender’s monitoring ability and incentives interact with the liquidation value of the firm and with its division among different claimholders. We propose a theory which reconcile the fact that private debt and bank debt are more protected in bankruptcy than public debt, even though the superior protection enjoyed by those creditors stifles their incentives to monitor. Our point is that different situations require different degrees of monitoring. Increasing the liquidation value for the monitors allows to have more monitoring in those states where monitoring is unconditionally beneficial without stifling managerial incentives.

The third chapter investigates how centralization and decentralization fare in imposing a hard budget constraint in a model with workers’ mobility across regions. In our model decentralization is associated with informational asymmetries among regions that make the cost of refinancing bad projects higher. Decentralization then implements a harder budget constraint when refinancing bad projects in one region creates a positive externality on other regions. The opposite holds in case of negative externality. The sign of the externality depends in our framework upon workers’ mobility costs across regions. The effectiveness of decentralization in implementing a hard budget constraint is enhanced by low mobility costs. Fiscal decentralization
should then be preceded by, or go together with, policies favoring mobility.

Thesis Supervisor: Oliver Hart
Title: Professor of Economics, Harvard University

Thesis Supervisor: Jean Tirole
Title: Professor of Economics
Acknowledgments

It is very hard for me to express how much I owe to Jean Tirole. I could say that without his support during the (many) confidence crises, I probably would have been unable to arrive until the end, but this would not be enough.

I could say that the ideas in this thesis owe so much to his ideas that I consider Jean my hidden co-author, but I am afraid that this would sound almost offensive for him.

I could say that he impressed me so much during these years that I will always try to be as generous, available, supportive with the students as he has been with me, but, again, this would not be enough.

I can only say that I was incredibly lucky to have Jean as an advisor. My only regret is that my thesis does not reflect at all the inputs he gave me.

I was also lucky to have Oliver Hart as an advisor. The discussions with him and his criticism have greatly improved my papers. He has always been very supportive even when there was no reason to be so. I have finally understood why he was so stubborn in asking me to relate my papers with the real world. I'm sorry it took me so long. Needless to say, I have towards Oliver the same regret I have towards Jean: his time could have been better spent.

My debt towards Mike Burkart, Francesco Daveri and Denis Gromb is almost too obvious to be acknowledged. They have been not only excellent co-authors, but great friends. This must have been hard given how aggressive I was during our discussions. A special thank to Denis: he had to suffer my bad mood many many times.

My gratitude also goes to the members of the ‘french-speaking’ network, i.e. Philippe Aghion, Patrick Bolton, Jacques Crémer and Mathias Dewatripont: I have greatly benefited from their comments and encouragement.

I did not have many friends among my colleagues at MIT, maybe because I liked (and I like) David Laibson, Tom Skinner and Jeromin Zettelmeyer so much. They have been great friends. The long hours spent with them have been by far the best moments in Cambridge for me.
I paid a very high price for this Ph.D.: I had to live without Vittoria for three years. If I had to do it again, I would strongly refuse. The price has been far too high. Thanks for all the love and care, Vittoria.

Dedication (in italian)

Questa tesi è dedicata ai miei genitori. Oggi che ho una figlia anche io, credo di capire il dolore che si prova a scoprire che la felicità e l'avvenire di un figlio passano per vie lontane e sconosciute ai genitori. Grazie per tutto l'affetto silenzioso e costante.
# Contents

1 Large Shareholders, Monitoring, and the Value of the Firm (joint with M. Burkart and D. Gromb) 8
   1.1 Introduction ........................................ 8
   1.2 Relation with the literature .......................... 11
   1.3 Initiative vs. monitoring. The static trade-off. ...... 14
   1.4 The dynamic trade-off ................................ 21
   1.5 Robustness to re trading .............................. 26
   1.6 Further applications .................................. 28
   1.7 Avoiding predation with a “deep pocket” ............. 30
   1.8 Productive conflict of interests ...................... 33
   1.9 Takeover threat and block premium ................... 37
   1.10 Comments ............................................. 40
   1.11 The impact of a fiduciary duty ....................... 42
      1.11.1 A simple model .................................. 43
   1.12 Extensions and future research ...................... 46
      1.12.1 The rigidity of ownership concentration ........ 46
      1.12.2 Debt structure and design ...................... 47
   1.13 Concluding Remarks .................................. 47

2 Debt Design, Liquidation Value, and Monitoring (joint with M. Burkart and D. Gromb) 49
   2.1 Introduction ........................................ 49
   2.2 The model .......................................... 53
2.3 Monitoring and the soft budget constraint  ..................  55
2.4 Monitoring and Liquidation Value  ..................  58
  2.4.1 A reinterpretation: management entrenchment  ....  64
  2.4.2 Recessions and booms  ..................  65
2.5 Multiple creditors  ..................  66
2.6 Different classes of claims: the role of seniority  ..........  72
2.7 Concluding Remarks  ..................  80

3 Decentralization, Mobility Costs, and the Soft Budget Constraint
  (joint with F. Daveri)  ..................  82
  3.1 Introduction  ..................  82
  3.2 Externalities and the soft budget constraint  ........  87
    3.2.1 The model  ..................  88
  3.3 Public good provision and migration as sources of externalities  ......  96
    3.3.1 The provision of local public goods  ..................  97
    3.3.2 Migration  ..................  98
  3.4 Externalities and mobility costs  ..................  101
  3.5 Mobility costs and the soft budget constraint  ..........  103
  3.6 Many regions  ..................  106
  3.7 Size of fiscal programs in centralization and decentralization  ..  109
  3.8 Concluding remarks  ..................  110

4 References  ..................  112
Chapter 1

Large Shareholders, Monitoring, and the Value of the Firm (joint with M. Burkart and D. Gromb)

1.1 Introduction

This paper provides a theory of the optimal ownership structure of firms by analyzing costs and benefits of ownership concentration, in particular the role of large shareholders. Empirically, large shareholders are an important feature of companies' ownership structure. Shleifer and Vishny (1986) report that more than half of the Fortune's 500 firms have at least one large shareholder holding a block exceeding 5%. The ownership structure does not only differ across firms but also evolves over time. Typically, young firms have a more concentrated ownership structure than more established ones. However, one also observes publicly held firms going private again or buying back their traded subsidiaries.

Our point is that the ownership structure of a firm may act as a commitment device to delegate a certain degree of authority from the shareholders to a self-interested management. More precisely, a dispersed ownership structure leaves relatively free hands to the manager while a more concentrated ownership results in closer moni-
toring of managerial activities, and hence less initiative. To the extent that initiative is valuable, we obtain a trade-off between the gains from monitoring and those from initiative, which is solved in the optimal ownership structure.

Starting with the seminal paper by Grossman and Hart (1980), the costs of dispersed ownership are quite well understood by the literature. The tendency of small shareholders to delegate authority originates in their free-rider behavior. Each shareholder’s stake is too small to incur the cost of supervising the management. Hence, in a widely held firm, the management is subject to little monitoring and hardly restricted in running the company at its own discretion. This is not without costs for the shareholders as the management’s and their interests do not always coincide. Typically, managers have also goals which are not shared by the shareholders, such as perquisites and power which they can pursue less impeded in the absence of close supervision.

Several authors have emphasized the role of large shareholders in disciplining the management and mitigating the free-rider problem. \(^{1}\). Due to her substantial stake in the firm, a large shareholder internalizes part of the externality. That is, the larger she is the more she is ready to provide the public good of monitoring the management and/or searching for a more profitable use of the corporate assets. Similarly, in Huddart (1993) and in Admati, Pfleiderer and Zechner (1994) the relation between the valuation of the firm and extent of monitoring is monotonically increasing in the size of the large shareholder’s block. Apart from the effects a blockholder has on the probability of a takeover attempt and the bid premium \(^{2}\), the presence of a large shareholder has no negative impact on the firm value. In a framework which does not rely upon the takeover mechanism, the only cost involved is usually the large shareholder’s private cost of holding an undiversified portfolio.

Thus, the argument offered to explain why we do not observe big firms owned by a

---

\(^{1}\)See e.g. Shleifer and Vishny (1986). Other devices to discipline the management in the pursuit of its self-interest are board of directors, competition on the managerial labor market an the product market, and the market for corporate control.

\(^{2}\)See e.g. Stulz (1988). For the large shareholder’s role in mitigating the free-rider problem that atomistic shareholders face when a single bidder with private benefits attempts a takeover, see Burkart (1993).
single party despite the increasing benefits from monitoring is the lack of diversification such a policy implies for the investor. This argument assumes that investors are risk-averse. While this assumption is appropriate for individual investors, it seems less convincing for institutional investors like investment banks or pension funds. Germany and Japan provide evidence of investment banks being large shareholders in some firms without owning all the shares. This, however, contradicts the above argument as it is clearly suboptimal for risk-neutral investors to limit themselves to a fraction of the companies’ shares.

Additionally, the trade-off between increasing benefits from monitoring versus higher costs of holding a less diversified portfolio implies that the larger the fraction held by the large shareholder the higher the value of the firm. However, some empirical studies, e.g. Morck, Shleifer and Vishny (1988), Wruck (1989), conclude that the relationship between the size of the shareholders’ block and the value of the firm is not trivial. That is, the relationship is not monotonically increasing, but rather piecewise linear, initially increasing, then decreasing and finally increasing again. In contrast, this paper provides an explanation why the presence of a large shareholder might be harmful to the value of the firm which is consistent with the observed non-monotonic relationship. As the underlying trade-off is between inducing managerial activism and maintaining control over the firm, our theory is complementary to the explanation where the initially increasing and then decreasing relationship between firm value and blockholder’s size results from the negative effect that managerial ownership has on the takeover threat.

After discussing the relation to the literature in the next section, section 3 illustrates in a static model the trade-off between monitoring and initiative. In section 4, we interpret this trade-off as one between monitoring and performance-based evaluation. In particular, we show that when the long-term profits are important, a large

---

3The argument that institutional investors are not risk-neutral because their managers are risk-averse agents does not invalidate this point. If institutional investors would indeed be risk-averse there would be gains from merging them. By increasing the fund’s size its risk exposure could be held constant while simultaneously increasing the fraction of shares held in each firm and thereby exploiting the benefits of increased monitoring.
shareholder may be desirable. In section 5, we show that our model is robust to retrading, i.e. that the optimal ownership structure is maintained even allowing for the possibility of selling or buying more shares. In section 6, we offer some further applications of our theory.

In section 7, we study the relationship between predation and ownership structure. If the signal on which the otherwise uninformed shareholders base their evaluation can be manipulated then the importance of monitoring increases and so does the optimal ownership concentration. In particular, when a rival firm can prey, the large shareholder may substitute performance-based evaluation with monitoring. In the case of predation this can be interpreted as a “deep-pocket” defense.

In part B the large and small shareholders have conflicting interests. In section 8 we show in the spirit of Crémer (1993b) how this can be exploited in order to promote managerial initiative. If the small shareholders are able to actually exert some formal authority they will not always believe the large shareholder’s report and sometimes overrule her. As a consequence the management is induced to try to produce a good signal in order to convince the small shareholders even if the large one learned his true type. In section 9 we show how the takeover threat makes it possible for the large shareholder to secure some private benefits. We then argue that this constitutes the basis of the block premium. In section 10 we comment on the previous sections 8 and 9. Section 11 discusses the impact of fiduciary duty when the large shareholder owns a majority block. We show how fiduciary duty may reduce her incentives to monitor. Section 12 presents extensions and directions for future research. Section 12 concludes.

1.2 Relation with the literature

Our approach is based on two recent contributions by Crémer (1993) and Aghion and Tirole (1994). Crémer shows that the principal may be better off committing not to acquire information about the quality of the agent. In such an arm’s length
relationship, the agent has to produce a signal proving he is of good quality. To do so, he needs to undertake a productive effort which benefits the principal. Aghion and Tirole show that the principal may be better off committing not to monitor the agent’s activity. The principal has *formal authority* in the sense that he has the power to overrule any decision taken by the agent. The problem is how to give the agent some *real authority*, i.e. the ability to take decisions which will not be reversed in the end. The idea is that the principal becomes more informed through monitoring and is more likely to overrule the self-interested agent. This in turn reduces the agent’s incentive to incur effort.

The principal’s ability to commit not to acquire information or not to interfere is an issue in both papers. In Crémer (1993), prior to the agent’s decision, the principal has to choose a monitoring technology. Aghion and Tirole (1994) suggest several commitment devices, such as a large span of control or splitting returns among multiple principals. By this means, the principal is committed not to monitor each of them too closely.

The present paper develops this theme in a corporate finance framework. The shareholders’ incentives to acquire costly information can be fine tuned through the allocation of the return rights. Hence, the ownership structure constitutes an ideal instrument to balance the gains from monitoring and those from managerial activism. In situations where initiative on part of the manager is especially valuable a dispersed ownership guarantees that the shareholders hardly overrule the manager. This in turn ensures that the manager does indeed work hard. Conversely, whenever monitoring is advantageous, a concentrated ownership structure induces the shareholders to undertake high levels of monitoring.

The papers closest to ours are Schmidt (1991) and Pagano and Röell (1994). Schmidt (1991) compares a nationalized and a regulated private firm. By assumption the government has more information about the firm’s costs and profits when the firm is nationalized. In his model, as in ours, the ownership structure can then be seen as a deliberate commitment device in order not to receive precise information
about the firm. In order to improve incentives for the manager, the government may want to commit ex-ante to a subsidy scheme which punishes the manager in case of bad performance. However, if the government is informed this commitment is not credible. In fact, the government will choose a production level which is ex-post efficient thus forgiving high costs and paying more subsidies than announced. Thus, under nationalization the government is unable to give the manager a high power incentive scheme because of the soft budget constraint. While Schmidt assumes that different ownership structures give rise to different levels of information, we endogenize this point, showing that a large shareholder acquires more information than the shareholders of a widely held public company.

In another related paper by Riordan (1990) on vertical integration information acquisition by the principal leads to a commitment problem similar to the one analyzed in Schmidt (1991). Under integration the principal, as the owner, has control of the accounting process and hence learns the actual cost parameter of the supplier/agent. As control over the accounting goes hand in hand with the ability to manipulate the accounts, the principal cannot use this information. That is, he cannot commit to a payment contingent upon the cost, as this would induce him to manipulate the accounts.

Pagano and Röell (1994) analyze the interactions between the design of ownership structure and the choice to stay private or go public. In their model too the role of the large shareholder is to monitor the management. The cost of excessive concentration is, as in our paper, overmonitoring. However they do not analyze the relationship between ownership concentration and managerial incentives.

Acemoglu (1994) also argues that there is a trade-off between the control benefits associated with tighter control (implemented by a concentrated ownership) and the costs of these in terms of worse incentives for the manager. In his model, as in ours, a disperse ownership structure creates a free-rider effect among shareholders and thus induce them to be passive. This mechanism allows managers to have discretion and hence better incentives.

Finally, Gertner, Scharfstein and Stein (1994) present a framework for analyzing
the costs and benefits of internal vs. external capital allocation. They isolate three factors to determine whether internal or external capital markets are preferable. Asset redeployability, hardness of budget constraint, and monitoring incentives. Their result that the public market works well when it is important to impose a hard budget constraint and monitoring is not a primary concern (and vice versa) is similar to our basic point.

The common theme in the above papers, as well as in Aghion and Tirole (1994) and Crémer (1993) is that information acquisition by the principal gives raise to adverse affects (in addition to the direct cost of obtaining the information). To resolve the trade-off between costs and benefits of information acquisition the principal has only two options at her disposal to choose from, e.g. to integrate or not in Aghion and Tirole, or to have a privately owned or publicly owned firm in Schmidt (1991). In contrast, while analyzing the same set of issues, this paper provides with the ownership structure a continuous rather than binary choice variable. Varying the size of the large shareholder’s block allows to balance the conflicting effects of information acquisition optimally.

PART A: MONITORING BY A LARGE SHAREHOLDER

1.3 Initiative vs. monitoring. The static trade-off.

The maintained assumption in this part of the paper is that the objective of the blockholder and those of the small shareholders are perfectly congruent, and more precisely that both classes of shareholders are only interested in the monetary returns. While this assumption might seem unrealistic, it allows us to present the basic trade-off associated with ownership concentration in a simpler way. The assumption is relaxed in part B where private benefits of the large shareholder are explicitly considered.
We want to illustrate that the management’s actual influence on the firm’s decisions increases as the size of the large shareholder decreases. This is due to two effects. First, the large shareholder’s incentive to acquire costly information depends positively upon the fraction of shares held by her. The smaller her block the more she, and all the other shareholders, have to rely upon the manager’s knowledge and judgement. Second, since this increases the value of his private knowledge the manager has greater incentives to invest in the information acquisition ex-ante as he decides the course of action more often. In the terminology of Aghion and Tirole (1994), he has more real authority the smaller the large shareholder’s block. That is, despite the fact that the shareholders have the formal authority many decisions are taken by the self-interested manager. As all shareholders have by assumption the same interests the allocation of authority among the shareholders is not an issue. The large shareholder’s actions are in the best interests of the small shareholders.

To show how a firm’s ownership structure determines the degree of authority delegated to the management we use Aghion and Tirole’s (1994) model and extend it with a richer structure on the side of the principal. Consider a firm in which a fraction \( \alpha \) is held by a large shareholder while \( (1 - \alpha) \) is dispersed among small shareholders. The risk-neutral owners of the firm have hired a manager to run the firm on their behalf. To keep things simple we assume that the manager is risk neutral as well and we abstract from monetary incentives \(^4\). That is, the manager’s compensation consists entirely of private benefits. His task is to search for a profitable project and propose it to the shareholders. The shareholders as the owner of the firm have the formal authority to accept his recommendation or to reject it and to pursue another course of action.

The firm must choose among many possible investment projects, \( i \in \{1, 2, \ldots, N\} \) which cannot be distinguished without further information. Abstracting from investment costs each project \( i \) when implemented yields verifiable security benefits \( \Pi^i \)

\(^4\)Monetary incentives are obviously a crucial determinant for the degree of congruence between the management’s and the shareholders’ objectives. Applying the same incentive scheme as in Aghion and Tirole (1994) would change our conclusions very little.
to all shareholders, and unverifiable private benefits $b^i$ to the manager. Thus, the
dlarge shareholder’s total payoff $B^i$ is equal to $\alpha \Pi^i$. However, only two projects are
potentially relevant, i.e. their two payoff components $B^i, b^i$ are positive, while for
all other $N - 2$ projects they are either zero or negative. It is convenient to assume
that one of these projects’ negative returns are so large that in the absence of further
information, all the parties prefer to stay inactive and get a zero payoff.

With probability $\lambda$, the interests of the management and those of the shareholders
are congruent, i.e. they get $b$ and $B$ when realizing one of the two potentially relevant
projects and 0 with the other. With the complementary probability $1 - \lambda$, their
interests are not congruent, i.e. one project pays $\{b, 0\}$ and the other $\{0, B\}$.

The manager can learn about the projects’ payoffs. By incurring effort $e$ at a
cost $c(e) = ce^2$ he perfectly learns the payoffs of all projects with probability $e$ and
with probability $(1 - e)$ he still cannot distinguish between the different projects.
The shareholders have access to a similiar binary technology. By incurring effort $E$
at a cost $k(E) = kE^2$ any of them can perfectly learn all the projects’ payoffs with
probability $E$ and nothing with probability $(1 - E)$. The information acquisition by
a shareholder is contemporaneous with the manager’s search effort $^5$.

Due to the free-rider problem the large shareholder is the only one who has an
incentive to provide (partially) the public good of being informed about the projects’
returns $^6$. However, her role is not confined to assist the manager in the search for a

$^5$Whether the (large) shareholder monitors simultaneously or after the agent has searched for a
profitable project does not affect our result. Only if the shareholder moves first and the outcome
of her monitoring effort is observable by the manager the result does not hold. In this case the
manager’s action is state-contingent. However, this order of moves can hardly be interpreted as a
delegation game.

$^6$In order to focus on the role of large shareholders auditing as an alternative means to monitor the
manager is ruled out. In our opinion there are several reasons which suggests that auditors are only
insufficient substitutes for active large shareholders. First, doubts about an auditor’s independence
towards the management are in place. Giving the board of directors the authority to choose the
auditor mitigates this problem but hardly removes all the management’s influence by virtue of the
latter’s influence on the board. Second, an auditor’s task is to inspect a company’s books rather
than to collect strategic information about future investment opportunities, as required in our model.
Third, even if auditors were to evaluate investment projects, in a world with incomplete information
supervision by a hired agent is unlikely to be equivalent to direct monitoring by the principal. To
ensure that hiring an auditor is worthwhile auditing models, e.g. Tirole (1986), Kofman and Lawarée
(1993), usually assume that the principal neither has the time nor the expertise to monitor, and
suitable project. As one of the many principals she also has the authority to overrule the manager's decision. Because $B^i = \alpha \Pi^i$, the small shareholders do not object when she is reversing the manager's project choice. She will use her power to overrule the manager's recommendation discretionary, that is, only if she has learned the identity of her preferred project.

Given that the large shareholder can overrule the manager and that their interests are not completely aligned the project will be chosen as follows. Whenever the manager knows the payoff structure he proposes to undertake the project he prefers most. If the manager does not learn anything he simply recommends not to invest. His recommendation is implemented whenever the large shareholder has no information of her own. Otherwise she reverses the manager's proposal with probability $1 - \lambda$ to realize her preferred project instead.

Hence, the manager's and the large shareholder’s optimization problems are as follows. Given $E$, the manager's utility is

$$U_M = E \lambda b + (1 - E)eb - \frac{c e^2}{2}$$

and the FOC gives

$$ce = b(1 - E) \quad (1.1)$$

Given $e$, the large shareholder’s utility is

$$V_L = EB + (1 - E)eB\lambda - k \frac{E^2}{2}$$

and the FOC gives

$$kE = B(1 - \lambda e) \quad (1.2)$$

Solving the system of equations (1.1) and (1.2) yields

---

that the auditor does not have the means to acquire the firm. However, an active large shareholder, as analyzed in this paper, conflicts with these assumptions as she is equivalent to a principal who has time and the expertise to monitor, or alternatively to an auditor who has the funds to acquire a large block in the firm.
\[
E = \frac{B(c - \lambda b)}{ck - \lambda bB}
\]
\[
e = \frac{b(k - B)}{ck - \lambda bB}
\]

As the small and large shareholders’ interests are by assumption aligned we can replace \( B = \alpha \Pi \) to get
\[
E = \frac{\alpha \Pi (c - \lambda b)}{ck - \alpha \lambda b \Pi}
\]
(1.3)

\[
e = \frac{b(c - \alpha \Pi)}{ck - \alpha \lambda b \Pi}
\]
(1.4)

**Lemma 1** For \( b < c \) and \( \Pi < k \), \( e \) and \( E \) ∈ (0, 1).

**Proof**

Follows directly from manipulation of equation (1.3) and (1.4). □

The equations above already illustrate the trade-off. It is easy to see that \( \frac{\partial E}{\partial \alpha} > 0 \) and \( \frac{\partial E}{\partial \alpha} < 0 \). That is, the greater the large shareholder’s block is the more effort she invests in monitoring. By equation (1.1) this in turn reduces the manager’s search effort as he is less likely to benefit from his private information.

Equation (1.3) characterizes the privately chosen effort level a large shareholder undertakes for a given block \( \alpha \). Due to the monotonic relationship between \( E \) and \( \alpha \) the optimal level of monitoring, i.e. the level which maximizes the firm value, can be implemented by choosing the large shareholder’s stake appropriately. The value of the firm is equal to the expected security benefits net of the direct monitoring cost \( k(E) = k \frac{E^2}{2} \). The expected security benefits is the sum of two elements. First, the return the shareholders get from their preferred project times the probability that the large shareholder identifies it, and second the expected return they earn when the manager’s preferred project is realized times the probability that only he learns the projects’ returns. Formally, the firm’s value is
\[ V_F = E\Pi + (1 - E)e\lambda\Pi - k\frac{E^2}{2} \]

**Proposition 1** When Lemma 1 holds and the interests diverge sufficiently, i.e. \( \lambda < \frac{e}{2b} \), the large shareholder’s optimal share is

\[ \alpha^* = \frac{k(e - 2\lambda b)}{ck - \lambda b(\Pi + k)} \]

**Proof**

Replacing \( e \) in \( V_F \) by its expression in \( E \) as given in equation (1.1) and differentiating with respect to \( E \) gives

\[ V'_F = \frac{\Pi}{c}(c - 2\lambda b) - \frac{E}{c}(ck - 2\lambda b\Pi) \]

\[ V''_F = -\frac{1}{c}(ck - 2\lambda b\Pi) \]

Suppose \( ck - 2\lambda b\Pi \leq 0 \). It then follows that \( c - 2\lambda b < 0 \). These conditions imply that \( V'_F \) is negative for any \( E \in (0, 1) \) and hence, \( \alpha^* = 0 \).

Suppose \( ck - 2\lambda b\Pi \geq 0 \) and \( c - 2\lambda b < 0 \). Then \( V'_F \) is negative for any \( E \in (0, 1) \) and hence, \( \alpha^* = 0 \).

Suppose \( ck - 2\lambda b\Pi \geq 0 \) and \( c - 2\lambda b > 0 \). Then \( V_F \) is concave and reaches its maximum at

\[ E^* = \frac{\Pi(c - 2\lambda b)}{ck - 2\lambda b\Pi} \]

Equating the privately chosen \( E \) with \( E^* \) and solving for \( \alpha \) yields the result. \( \square \)

We assume that there is only one large shareholder who monitors and many small shareholders who free-ride. Obviously, any given level of monitoring induced by the large shareholder’s block \( \alpha \) could also be implemented by two or more large shareholders. However, the model is implicitly geared towards one large shareholder who is not meant to be just one person. The underlying intuition is that one party can always do at least as well as two and most likely better by copying what they are
doing while avoiding any duplication. Additionally, the implicit assumption of only one large shareholder is in accordance with empirical findings. Zwiebel (1991) reports that in firms with large blockholders there is typically only one party owning a substantial large block, i.e. firms with multiple large blockholders are rare. When sticking closely to the model, the question of whether one or more large shareholders are optimal depends upon the trade-off between the duplication of monitoring effort vs. the convexity of monitoring costs. For instance, if there is perfect correlation between the monitoring efforts $E_1$ and $E_2$ incurred by shareholder 1 and 2, one large shareholder will always be optimal. If there is only partial duplication, which occurs even when $E_1$ and $E_2$ are independent, the optimality of one monitor requires that the cost of duplication outweighs the increase in the marginal monitoring cost. The duplication of monitoring efforts by two (or more) large shareholders implies that the sum of their stakes is always greater than the size of the block a single large shareholder needs to hold in order to implement the same level of supervision (or respectively the same level of managerial initiative) \(^7\). Another aspect which needs to be considered when allowing for more than one large shareholder is the strategic interaction among monitoring parties. If one assumes that large shareholders behave non-cooperatively, multiple monitors faces a prisoner's dilemma situation as each individual blockholders attempts to free-ride on other's monitoring effort \(^8\). This effect obviously increases further the aggregated blocks two (or more) large shareholders need to own in order to implement a given level of supervision.

**Proposition 2** The optimal fraction of shares held by the large shareholder is decreasing in the congruence parameter $\lambda$.

\(^7\)To see this consider the case of two large shareholders where their monitoring efforts are independent. Due to the convexity of the monitoring costs two symmetric large shareholders, i.e. both owning a block of identical size, is clearly optimal. To implement any given level of managerial initiative w.l.o.g. two shareholders instead of one, it must hold that $(1 - E^1) = (1 - E^2) \in (0, 1)$ where $E^1$ is the monitoring effort of a single large shareholder and $E_i, i = 1, 2$ is the effort by one of the two shareholders. From this equality follows that $E^2 = E^1 + (E^1)^2 > E^1 \cdot \alpha(E^1) > 0, E^2 > E^1$ implies that $2\alpha(E^2) > \alpha(E^1)$.

\(^8\)For examples of this point, see Crémer (1994) and Povel (1994).
Proof Simple computations show that

$$\frac{\partial \alpha^*}{\partial \lambda} = \frac{ckb(\Pi - k)}{(ck - \lambda b(\Pi + k))^2}$$

By the conditions of Lemma 1 the numerator is negative. \qed

This result perfectly matches the one obtained by Aghion and Tirole (1994) when comparing integration vs. non integration. While the former gives both formal and real authority to the principal, the latter is a means to commit to give the agent real authority. They show that non integration is preferable to integration when the congruence of interests between manager and principal is high. Our result that the optimal degree of concentration of ownership is decreasing with the congruence of interests can be read as the continuous version of this result. When the manager can be trusted to take decisions in the shareholders' interests it is better to leave him discretion in order to promote initiative. This is achieved by allocating only a small block to the large shareholder. The opposite holds when there is little congruence of interests.

1.4 The dynamic trade-off

In this section we want to analyze how the ownership structure can be used as a device to balance the conflicting effects to which the information acquisition by shareholders can lead in a slightly different setting. The trade-off involved is as follows. On the one hand a large shareholder has incentives to monitoring the manager closely. Thereby she reduces the risk that a bad manager retains his position and avoids low returns in the future. On the other close monitoring reduces the manager's current effort. The more the large shareholder monitors the higher the chances that a bad manager is identified as such despite having produced high current profits. Conversely, in a widely held firm without any monitoring the decision to retain the manager is solely based upon current performance, i.e. this period's profits. This induces the manager to work hard, boosting current profits at the cost of retaining bad managers.
The trade-off in section 1.3 is between eliciting effort from the manager and curbing his pursuit of private interests, and the optimal ownership structure depends on the congruence. In the initial model, the degree of congruence is just taken as given. In the dynamic framework we allow for a varying extent of congruence of interests. At the beginning the firm is locked in with the incumbent management team. But in the future the shareholders might have to decide whether to allow the incumbent management to remain in control. The relative size of the short-run profits with respect to long-term profits is a measure of congruence of interests. We will show, in fact, that the optimal stake of a large shareholder is inversely related to the importance of short-term profits. On the other hand, it is positively related to the importance of long term profits.

To show how the ownership structure can influence the effort as well as the quality of the managers we use Crémer’s (1993) model with some minor variations. Specifically, instead of choosing between a binary monitoring technology we vary the extent of ownership concentration which in turn gives rise to varying degree of monitoring. Consider again a firm in which a fraction \( \alpha \) of the shares is held by a large shareholder while the remaining \( 1 - \alpha \) is dispersed among small shareholders. All shareholders are risk-neutral and their interests are congruent. In contrast to the previous section the firm is assumed to live for two periods. In the first period its profits can take two values, \( \Pi_1 \) or 0. In the second period they are either equal to \( \Pi_2 \) or to 0, where \( \Pi_1 > 0 \) and \( \Pi_2 > 0 \). \( \Pi_1 \) and \( \Pi_2 \) are verifiable.

At the beginning of the first period the shareholders hire a manager from the labor market to run the firm on their behalf. For simplicity we assume that the manager has only private (non-monetary) benefits \( b \) from remaining in control in the second period. All managers are risk-neutral and can be of two types, either good or bad, which reflects their long-term performance. In each period, the manager can

---

One may wonder why is not preferable that the ownership structure evolves over time. We remind the reader that while reducing the concentration of ownership is without costs, increasing the concentration might be extremely costly, due to the free-rider problem of Grossman-Hart (1980). So the firm might be obliged to choose initially an allocation of ownership that reflects the future need for a higher degree of concentration.
produce an effort \( e_i \in [0, 1] \) at cost \( c(e_i) = c_i^2 / 2 \). A good manager yields profits \( \Pi_1 \) in period 1 and \( \Pi_2 \) in period 2, irrespective of his effort. A bad manager produces \( \Pi_i \) with probability \( e_i \) and 0 with probability \( 1 - e_i \). The prior probability that the manager is good is \( p \). As nobody knows the type of the manager, not even himself, type refers to the manager's suitability for the job or to the quality of the project.

All shareholders have access to a binary information acquisition technology. By incurring effort \( E \in [0, 1] \) at cost \( k(E) = kE^2 / 2 \) any of them can learn the manager's type with probability \( E \) and learns nothing with probability \( 1 - E \). As none of the small shareholders has any incentive to do so only the large shareholder will incur the search costs in order to identify the manager's type. If she finds out that the manager is bad, she can fire him and employ another one from the labor market. The probability that a good manager is available on the labor market is again \( p \). The manager hired in the first period is replaced in two cases. Either there is checking and the manager is found to be of bad quality or the first period profit produced under his management is 0.

Notice first that \( e_2 = 0 \) as the only incentive the manager has to incur effort is to remain in control. Hence, only the quality of the management team in charge is relevant for the second period profits. This is only a convenient simplifying assump-

---

10 The reader might wonder why we have decided to opt for this specific stochastic structure and not for a more general one. For instance, we could assume that the able manager has a probability \( p(e_i) \) of obtaining \( \Pi_1 \), given effort \( e_i \), while the unable manager has probability \( q(e_i) \), with \( p(e_i) > q(e_i) \). However, the crucial variable is simply the relationship between \( p'(e_i) \) and \( q'(e_i) \). In particular, if \( p'(e_i) > q'(e_i) \) for all \( e_i \), then we always obtain a bang-bang solution, i.e. the optimal share of the large shareholder is either 0 (or better, a fraction of shares close to 0) or 1, i.e. a private firm. In order to have an interior solution we need the assumption \( p'(e_i) < q'(e_i) \) at least for some values of \( e_i \). The specification chosen in this section obeys this last inequality and it allows us to obtain an explicit solution for \( \alpha^* \).

11 In the present model offering severance payments to identify a bad manager is not a feasible alternative to monitoring as the managers do not know their type. However, even if they would know their type severance payments might be too costly to be profitable. In order to induce a bad type to accept it the shareholders have to compensate him for the perks he expects to get if he would take the employment offer instead. Additionally, there are considerable costs involved in accepting severance payments as the manager might reveal his type to outsiders and thereby jeopardizing his career prospects severely. Severance payments have to cover these costs in addition to the forgone perks. Furthermore, even in a setting where severance payments are worthwhile, monitoring as an instrument to uncover bad types cannot be entirely dismissed. For example, it could be used in order to reduce the level of severance payments offered.
tion which does not affect the qualitative results. Now we denote without ambiguity the first period effort \( e_1 = \epsilon \). Hence, the manager's and the large shareholder's optimization problem are as follows. Given \( E \), the manager's expected utility is

\[
U_M = b[p + (1 - p)e(1 - E)] - c\frac{e^2}{2}
\]

and the FOC gives

\[
ce = b(1 - p)(1 - E)
\] (1.5)

Given \( e \), the value of the large shareholder's block is

\[
V_L = \alpha \{p(\Pi_1 + \Pi_2) + (1 - p)e[E(\Pi_1 + p\Pi_2) + (1 - E)\Pi_1] + (1 - p)(1 - e)p\Pi_2\} - k\frac{E^2}{2}
\] (1.6)

and the FOC gives

\[
kE = \alpha p(1 - p)\Pi_2 e
\] (1.7)

Thus, the monitoring is higher the larger the fraction held by the large shareholder, the higher the uncertainty about the manager's type, and the higher long-term profits. By equation (1.5) more monitoring affects the level of effort adversely. Close checking increases the risk of being identified as a bad manager and hence reduces the incentive to produce the costly signal \( \Pi_1 \). Solving the system of equation (1.5) and (1.7) yields

\[
e = \frac{b(1 - p)k}{ck + \alpha p(1 - p)^2 b\Pi_2}
\] (1.8)

\[
E = \frac{\alpha p(1 - p)^2 b\Pi_2}{ck + \alpha p(1 - p)^2 b\Pi_2}
\] (1.9)

Under the condition \( b(1 - p) < c \) we have an interior solution. Equations (1.8) and (1.9) characterize the privately chosen levels of effort and monitoring. The optimal level of monitoring, i.e. the level which maximizes the verifiable security benefits net of the monitoring costs, can again be implemented by choosing the size of the large shareholder's stake appropriately. The value of the firm is
\[ V_F = p(\Pi_1 + \Pi_2) + (1-p)e[E(\Pi_1 + p\Pi_2) + (1-E)\Pi_1] + (1-p)(1-e)p\Pi_2 - \frac{E^2}{2} \]  

(1.10)

Proposition 3  The optimal fraction of shares held by the large shareholder is

(a) \( \alpha^* = 0 \) if \( \Pi_1 > 2p\Pi_2 \)

(b) \( \alpha^* = 1 \) if \( \Pi_1 < p\Pi_2[1 + \frac{(1-p)^2b\Pi_1}{ck}] \)

(c) \( \alpha^* = \frac{2p\Pi_2 - \Pi_1}{p\Pi_2[1 + \frac{(1-p)^2b\Pi_1}{ck}]} \) otherwise

Proof

By differentiating \( V_F \) with respect to \( \alpha \) we get

\[
(1-p)\Pi_1 \frac{\partial e}{\partial \alpha} + p(1-p)\Pi_2[e \frac{\partial E}{\partial \alpha} - \frac{\partial e}{\partial \alpha}(1 - \lambda) - kE \frac{\partial E}{\partial \alpha}]
\]

Define \( D = ck + \alpha p(1-p)^2b\Pi_2 \), so that \( \frac{\partial E}{\partial \alpha} = \frac{ckp(1-p)^2b\Pi_2}{D^2} \) and \( \frac{\partial e}{\partial \alpha} = -\frac{k\alpha p(1-p)^2b^2\Pi_2}{D^2} \)

Substituting and after some algebra we obtain

\[
D^3 \frac{dV_F}{d\alpha} = [p\Pi_2 - \Pi_1]D + p\Pi_2[(1 - \alpha)ck - \alpha p(1-p)^2b\Pi_2] = 0
\]

which can be further simplified to

\[
2p\Pi_2ck - \Pi_1ck - \alpha p\Pi_2[ck + (1-p)^2b\Pi_1] = 0
\]

from which we derive the final result. It is also clear that the interior solution is a maximum.

\[ \square \]

Corollary 1  The optimal ownership structure is more concentrated (\( \alpha^* \) increases) as long-term profits become more important (\( \Pi_2 \) increases) and/or short-term profits become less important (\( \Pi_1 \) decreases).

Proof

To see that \( \alpha^* \) is decreasing in \( \Pi_1 \) just observe that \( \Pi_1 \) enters in the expression of \( \alpha^* \) with a positive sign at the denominator and with a negative sign at the numerator.
To see that it depends positively on $\Pi_2$ rewrite the expression as
\[
\frac{2p - \frac{\Pi_1}{\Pi_2}}{\frac{2p - \Pi_1}{\frac{1}{2}(1 - \rho)^2\frac{\Pi_1}{\Pi_2} \frac{1}{ck}}}
\]
which makes the result obvious.

Notice that we have proved the result that a higher congruence of interests i.e. $\Pi_1$, reduces the optimal fraction of shares owns by the large shareholder, confirming the results of the previous section.

### 1.5 Robustness to retraining

In the section above we have found a theory of optimal ownership structure. Obviously, a very important question is whether the optimal ownership structure is robust to retraining. In other words, suppose that the large shareholder could sell (part of) his shares before the monitoring effort $E$ has been chosen. Has he an incentive to do so? The answer is a qualified no. We prove it only for the model of section 1.3. The proof for the dynamic case runs along the same lines.

**Proposition 4**

The large shareholder has no incentive to alter his stake from $\alpha^*$ by subsequent trading if all investors are rational and have perfect knowledge of the size of the trades and of the identity of the traders.

**Proof** Equations 1.6 and 1.10 can be written as

\[
V_F(\alpha) = \Pi(E(\alpha)) - k[E(\alpha)]^2/2
\]

\[
V_L(\alpha) = \alpha \Pi(E(\alpha)) - k[E(\alpha)]^2.
\]

The maximization of firm value is equivalent to

\[
\max_E \Pi(E(\alpha)) - k[E(\alpha)]^2/2
\]

---

12Our Proposition parallels Proposition 3 in Pagano and Röell (1994).
that gives
\[
\frac{\partial \Pi(E(\alpha))}{\partial E} - \frac{k \partial [E(\alpha)]^2}{2 \partial E} = 0.
\]

Let \( E^* \) be the solution to the problem above.

Maximizing the value of the firm requires to set \( E^* = E(\alpha^*) \), taking into account that \( E(\alpha) \) is privately chosen by the large shareholder to maximize
\[
V_L(\alpha) = \alpha \Pi(E(\alpha)) - k[E(\alpha)]^2/2.
\]

From the first order condition of the large shareholder we get
\[
\alpha^* \frac{\partial \Pi(E(\alpha^*))}{\partial E} - \frac{k \partial [E(\alpha^*)]^2}{2 \partial E} = 0 \tag{1.11}
\]

Suppose the large shareholder wants to change his initial stake \( \alpha^* \) by a fraction \( \Delta \) of the firm's shares. Then the stock price set by rational and farsighted investors will be \( V_F(E(\alpha^* + \Delta)) \).

The large shareholder's return from selling a fraction \( \Delta \) of shares is
\[
(\alpha^* - \Delta) \Pi(E(\alpha^* - \Delta)) - \frac{k[E(\alpha^* - \Delta)]^2}{2} + \Delta \Pi(E(\alpha^* - \Delta))
\]

which can be rewritten as
\[
\max_{\Delta} \alpha^* \Pi(E(\alpha^* - \Delta)) - \frac{k[E(\alpha^* - \Delta)]^2}{2}.
\]

The first order condition with respect to \( \Delta \) for this problem is
\[
-E'(\alpha^* - \Delta) [\alpha^* \frac{\partial \Pi((E(\alpha^* - \Delta))}{\partial E} - \frac{k \partial [E(\alpha^* - \Delta)]^2}{2 \partial E}] = 0
\]
This condition is always satisfied for \( \Delta = 0 \) since the expression in brackets is zero at \( \alpha = \alpha^* \) by equation 1.11.

Obviously any trade of the small shareholders among themselves does not change the value of the firm. Hence our model is robust to retrading.
1.6 Further applications

- Another way of thinking of effort is to view it as a firm-specific human capital investment on the part of the management. For example, suppose that by investing the manager could reduce the search cost \( c(e) \). However, a manager will invest in the acquisition of firm-specific knowledge only if he can recoup his investment cost. His prospects of doing so are determined by the level of monitoring the firm’s ownership structure induces. Close monitoring reduces the extent of his real authority or alternatively increases his risk of being identified as a bad manager. Hence, close checking lessens the incentive to invest. Firm-specific investment is best promoted through a widely dispersed ownership. Furthermore, a manager who has already invested in firm-specific knowledge faces a potential hold-up problem. If the ownership structure should become more concentrated in the future he will be more closely monitored and might hence be unable to capture sufficient benefits to make the investment ex-post profitable.

- The result that a large shareholder is good when long-term profits are important, but is bad when initiative is important has some interesting implications for privatization policy. In particular, it implies that there is no unique answer to the privatization debate going on currently in Italy about the optimal ownership structure of firms which must be privatized. More precisely, we believe that firms which are to be privatized should aim for different types of ownership structures. For example, firms with low current profits which need to be reorganized to achieve profits in the future should opt for a concentrated ownership, i.e. a large shareholder who takes an active role in the restructuring of the company. On the contrary, for firms that are basically in good shape a highly dispersed ownership ensures that no undesirable interferences curb the management initiative\(^\text{13}\). Thus, contingent on the condition of the company the optimal ownership structure for privatization may either be a public company or a half public company where the state or another investor retains a large

\(^{13}\) Again, we make use of the argument, borrowed by Grossman and Hart (1980), that once the ownership structure is dispersed, it is costly to concentrate it, due to the free-rider problem.
block.

- The conventional wisdom on going public/private is the so-called ageing theory. At the beginning, firms are privately held. When they expand, the founder has to go public due to insufficient resources of his own to finance the growth. Zingales (1992) offers another possible view, arguing that for entrepreneurs who are willing to sell their firm, going public is a means to extract (most of) the surplus, by using the free-rider behavior of the small shareholders. Pagano and Röell (1994) argue that a firm decides to go public in order to reduce liquidity cost, paying however the price of a high listing cost.

Our approach suggests further factors which influence the decision to go public/private. Firms are private at the beginning of their life and in periods of big changes, i.e. whenever monitoring is the important input. Once a firm has established a good reputation, monitoring becomes less important and initiative is emphasised. That is, a large shareholder is valuable in periods in which it is important to match the factors of production in an efficient way. After this task is realized, a large shareholder can have the tendency to excessively interfere with the management activity, and thereby stifling initiative.

- Numerous academics 14, the press and the practitioners have discussed the alleged short-termism of the Anglo-Saxon economies and contrasted it with its absence in economies such as Germany or Japan. 15 This difference is usually attributed to the interaction of financial markets with managerial decision making. That is, the prominence of market-based performance evaluation induces the Anglo-Saxon manager to emphasize the short-term while the presence of large investors, typically banks, allows firms in Germany or Japan to take a long-term perspective. By modifying our dynamic model slightly it might capture, although rather crudely, the spirit of this explanation. Suppose that current effort also affects long-term profits. In a widely

---

14 See e.g. Dewatripont and Maskin (1995)
15 Porter (1992)'s empirical study provides evidence of short-termism in the U.S. Miles (1993), analysing stock market data of a large sample of U.K. companies, finds evidence that longer-term discount rates are “too high”, a result consistent with short-termism.
held firm the evaluation of the manager is solely based upon current performance. The current profit’s signalling function induces a manager to overvalue the short-term aspects of his actions. Hence in an economy dominated by dispersed ownership managers are confronted with this incentive structure, and short-termism will appear. In contrast, in a firm with a large shareholder the decision to retain or fire the manager depends upon current profits as well as the information gathered through monitoring. This reduces the current profit’s signalling value and hence the incentives to engage in short-term actions for the purpose of transmitting one’s quality to the market. Thus, in economies characterized by active large investors managers can adopt a more balanced view of the firm’s short- and long-term performance.

1.7 Avoiding predation with a “deep pocket”

In this section we offer a new perspective on the “deep pocket” argument by analyzing the relationship between predation in the product market and ownership structure. Bolton and Scharfstein (1990) show that the contract which minimizes agency costs between the lender and the borrower when the cash-flow in unverifiable simultaneously maximizes the incentives to prey on part of a product market rival. The reason is that the agency-cost minimizing contract maximizes the difference in the probability of being refinanced in the second period when the borrower makes a repayment and when he does not. There are two possible ways to amend the contract to the threat of predation. First, a “deep pocket” strategy where the firm is refinanced with positive probability even if there is no repayment, and second a “shallow pocket” strategy where the plug is pulled with positive probability even if there is repayment. Bolton and Scharfstein (1990) show that the optimal contract is the shallow pocket strategy.

We try to argue instead that in our framework a situation closer to the spirit of the deep pocket strategy emerges. Contrary to Bolton and Scharfstein’s model

---

16Stein (1988) has first formalized the idea that a market-based system might force managers to engage in wasteful signalling. In his model the takeover threat induces managers to sacrifice long-term profits in order to boost current earnings. This avoids mispricing by the market and thereby prevents a raider from taking the firm over at an undervalued price.
where the cost of observing the cash flow is infinite for the creditor by assumption, we argue that the cost of distinguishing between slack and predation depends on the ownership structure of the firm. As we have shown in the previous sections, a more concentrated ownership structure leads to higher levels of monitoring and hence reduces the incentive for a rival to prey on the firm. Our version of the deep pocket strategy is as follows. In order to reduce the threat of predation, a firm will tend to bias its ownership structure towards more concentration, thereby inducing more monitoring and thus reducing the incentives to prey.

The model is a simple variation of the one presented in section 1.4. The only modification is the reversed stochastic structure. Now the good type can obtain a high signal in the first period with probability \( e \) at a disutility \( \frac{e^2}{2} \) and a low signal with the complementary probability. The bad type obtains the low signal for any level of effort. It is easy to check that the optimal ownership structure in this framework is as follows. \( \alpha^* = 0 \) if \( b \geq \hat{b} \), or \( \alpha^* = 1 \) if \( b \leq \hat{b} \), where \( \hat{b} \) is a cut off point which depends on the parameters of the model. That is, we have a bang-bang solution. Either a public company with dispersed ownership or a fully concentrated ownership are optimal. In particular, higher incentive to retain control leads towards a more dispersed ownership.

We model the possibility of predation by assuming that the rival can decrease by \( \mu \) the probability of a high signal at a cost \( \frac{\mu^2}{2} \). If the good type remains in control, the firm realizes a profit \( \bar{\Pi} > 0 \) and the rival a profit \( R \). We further assume that it is never optimal to have a low ability manager in control, so that the firm stops to produce and a liquidation value \( L < \bar{\Pi} \) is paid. If the firm is closed down the rival has a profit \( R^M > R \).

Hence, the manager's, the rival's and the large shareholder's payoff functions are

\[
U_m = pb\left(e - \mu + (1 - e + \mu)E\right) - \frac{e^2}{2}
\]
\[ U_R = p\{[e - \mu + (1 - e + \mu)E]R + (1 - e + \mu)(1 - E)R^M\} + (1 - p)R^M - \frac{\mu^2}{2} \]

\[ V_L = \alpha\{p[e - \mu + (1 - e + \mu)E]\tilde{\Pi} + p(1 - e + \mu)(1 - E)L + (1 - p)L\} - \frac{E^3}{2} \]

and the FOCs are

\[ e = pb(1 - E) \]

\[ \mu = p(1 - E)(R^M - R) \]

\[ E = \alpha p(1 - e + \mu)(\tilde{\Pi} - L) \]

We can define two new variables \( \tilde{e} = e - \mu \) and \( \tilde{b} = b - R^M + R \) to rewrite so the conditions as

\[ \tilde{e} = p(1 - E)\tilde{b} \]

\[ E = \alpha p(1 - \tilde{e})(\tilde{\Pi} - L) \]

Hence, the model is the same as before with the exception of the redefined variables. Notice in particular that \( \tilde{b} < b \) so that predation is equivalent to a reduction in the manager's incentive to work. We already know that this leads to a more concentrated ownership structure, vindicating the "deep pocket" conjecture.
PART B: CONFLICTING INTERESTS

In this part of the paper we address the case in which the large shareholder enjoys benefits other than security benefits. That is, her share in the firm allows her to receive private gains that do not accrue to the other shareholders. This can lead to a possible conflict of interests between the small and the large shareholders.

The fact that a large shareholder may enjoy private gains is quite well documented. In the absence of private gains, blocks of shares ought to be sold at a discount due to the greater risk exposure and due to the monitoring costs. However, blocks are usually sold at a premium which suggests the presence of private gains. Private gains may be pecuniary or non-pecuniary, and may stem from taking decisions which actually reduce the security benefits, e.g. use the firm's structures for personal purposes, or engage into sweetheart deals.

The case of Koito and Toyota provides an example of a large shareholder's enjoying private benefits at the expense of the other shareholders and being stopped. In 1989 the raider T. Boone Pickens alleged that Koito, a Japanese auto part maker, had for many years been subsidizing sales to its large shareholder, the Toyota Car Group. He bought a block and demanded board representation in order to stop the diversion of corporate resources\textsuperscript{17}. Although this is surely an extreme case, private gains will undoubtedly have an impact on the decisions of a large shareholder.

1.8 Productive conflict of interests

When the interests of the large and small shareholders are potentially conflicting, contrary to the previous sections, the allocation of authority among the shareholders becomes an issue.

We consider in this section the case in which the large shareholder holds a minority block. In this case, the large shareholder cannot directly implement her preferred

\textsuperscript{17}Reported in Barclay and Holderness (1989).
decisions, but she needs the consent of the general assembly or the approval of the board, whose members are nominated by the shareholders.

The main idea of this section is that the potential for conflicts among shareholders may be productive in the sense that it allows the manager more initiative and/or provides him with better incentives. Our point is that a limited stake of the large shareholder is a commitment not to always (be able to) use all of the acquired information. Because of the potential clashes of interests, the majority of the shareholders does not fully trust her reports and can overrule her decisions, although sometimes wrongly. The degree of the large shareholder’s real authority increases as her interests becomes more aligned with the other shareholders’s interests\(^{18}\). But more authority for the large shareholder means less authority and initiative for the management. Hence, the conflict of interests and the limited alignment are a commitment to judge the management on his (publicly) observable performance, rather than on the result of monitoring. This in turn, as in section 1.4 increases the management’s incentives to produce good results.

To formalize this point, we use the variation of the dynamic model we proposed in section 1.4, with the following four additional assumptions.

**Assumption 1** The large shareholder derives private benefits \(Z\) if and only if the initial manager is retained.

This assumption creates a contrast between large and small shareholders. The large shareholder’s interest is a combination of that of the small shareholders and of that of the management\(^{19}\).

\(^{18}\)This assumption is commonly made in the corporate finance literature and rules out different attitudes towards risk as the source of the conflict of interests. Typically, the large shareholder’s portfolio will be less diversified than the ones the small shareholders hold. Hence, she will tend to favour safer projects than the other shareholders. This difference tends to increase rather than decrease as the stake of the large shareholder increases.

\(^{19}\)Though we do not discuss this point here, it could probably be endogenized as the result of collusion between the management and the large shareholder. If we reinterpret the manager’s type as the quality of his projects, we could tell the following story. The current projects were chosen by the management so as to please the large shareholder. By now, the management has invested in human capital specific to these projects. Hence, for both the management and the large shareholder, terminating the project is costly.
In order to be able to talk about the limited trust that small shareholders credit the large shareholder's reports with, the information these contain should not be enough to convince them.

**Assumption 2** The information the large shareholder acquires is soft in the sense that it cannot be credibly conveyed to other parties.

**Assumption 3** The decision whether to maintain or replace the management must be approved by simple majority which is in the hands of the small shareholders.

In practice, the large shareholder can also be overruled by the board of directors, in proxy fights or takeovers. More generally, the reactions of the financial market to some corporate decisions may matter. (In a more general model, the probability of the large shareholder being overruled should also depend on her size.) Obviously, the assumption keeps the analysis simple.

**Assumption 4** Monitoring produces additional security benefits ET.

The need for this assumption will become clearer later on. We have seen a way of interpreting this assumption in the static model. The large shareholder can find herself the profitable project even if the manager does not incur any effort. In the dynamic framework it can be interpreted as follows. There is a third type of managers, very bad ones. They occur with a very low probability but produce extremely low security and private benefits (like the N-2 projects of the static model). Hence, there is no conflict of interest between the large and small shareholders concerning their dismissal. An alternative interpretation is simply that monitoring is needed and useful in some other activity of the firm.

Depending on the size of her stake, the interest of the large shareholder is more or less aligned with that of the small ones. In this very simple model there are only two cases to distinguish.

---

20In a more general setting the private benefits of the large shareholder Z is a random variable and the probability of aligned interests is increasing in the stake of the large shareholder.
Aligned interests

When the stake of the large shareholder is big enough, the private gains she would secure by keeping a bad manager fall short of the costs. Hence, she will report truthfully the manager's type, and the small shareholders believe her.

If $\alpha p \Pi_2 \geq Z$, there is full alignment of interests. Hence, we are essentially back to the analysis of section 1.4, (we use the index $a$ for “aligned”, and $d$ for “diverging”).

$$I'_M = b[p + (1 - p)e_a(1 - E_a)] - c \frac{e_a^2}{2}$$

$$V_L = \alpha E_a T + p[\alpha(\Pi_1 + \Pi_2) + Z] + (1 - p)[e_a(\alpha \Pi_1 + Z) + e_a E_a(\alpha p \Pi_2 - Z)] - k \frac{E_a^2}{2}$$

and the FOCs are

$$ce_a = b(1 - p)(1 - E_a)$$

$$kE_a = \alpha T + (1 - p)(\alpha p \Pi_2 - Z)e_a$$

Diverging interests

In contrast, if the large shareholder's stake is small enough, then she has an interest to promote even bad managers. The private gains exceed the loss in his limited share of the security benefits. As a consequence the small shareholders do not trust the large shareholder's recommendation and they base their decision whether to retain or replace the manager solely on the first period profit.

If $\alpha p \Pi_2 \leq Z$, the objectives differ: the large shareholder would like to retain bad and good managers. The manager maximizes

$$U_M = B[p + (1 - p)e_d] - c(e_d)$$
\[ c e_d = B(1 - p) \]

The large shareholder's monitoring has no impact on the decision to continue or stop the project. Only the production of the additional security benefits motivates her monitoring efforts. Here assumption 4 is needed. Without it, monitoring would be of no value in this two-type model. Hence, conflicting interests refer to a situation where the interests of the large and the small shareholders are different but not completely antagonistic. The large shareholder incurs monitoring effort such that

\[ k E_d = \alpha T \]

In this simple two-type model, the conflict of interests makes \( e \) and \( E \) independent of each other. Note that for \( e_d \) to match \( e_a \), \( E_a = 0 \) would be required. For instance, if monitoring costs are extremely low but still positive the management would not make any effort at all. The conflict of interests created by a "small" large shareholder prevents this extreme moral hazard problem. Thus, when the manager's effort is valuable for the shareholders, it can be optimal to have a relatively small large shareholder so as to generate this productive conflict of interest.

As a by-product of the model, we illustrate that the large shareholder's private gains are a function of her size. However, when the size of her stake varies two conflicting effects are at work and the net effect is a priori ambiguous. On the one hand, as \( \alpha \) increases the large shareholder is more and more trusted by the others. This enables her to impose her preferred option and to secure private benefits more often. On the other hand, her preferred option coincide increasingly more often with the one of the small shareholders and so she pursue less and less her private gains.

### 1.9 Takeover threat and block premium

In this section we show how the large shareholder's ability to threaten a takeover forms the basis of her private gains. In order to avoid a takeover the manager takes
the large shareholder’s view into account when recommending an investment project.

To formalize the takeover threat we adapt the static model. If the manager exerts effort \( e \) he finds a project with probability \( e \) and with the complementary probability \( (1 - e) \) his search efforts are in vain. A project when implemented yields security benefits \( \tau \Pi \) with \( \tau \in \mathbb{R} \), no private benefits to the large shareholder, and private gains \( b \) to the manager. The quality \( \tau \) of the project is a random variable distributed on \( \mathbb{R} \) according to the distribution function \( f \), with c.d.f. \( F \). \( F \) is twice continuously differentiable. Both the manager and the large shareholder observe \( \tau \) with certainty if the former finds the project.

For simplicity, we assume that the large shareholder does not search for projects. However the large shareholder has always access to an outside option which yields private gains \( Z \) to her and zero returns to the manager and small shareholders.

The large shareholder can takeover the firm in which case he receives the full formal authority. We assume that the large shareholder has to incur takeover costs \( k(\alpha) \), which are decreasing in \( \alpha \).

The manager’s project is implemented when the large shareholder does not have an incentive to take control, i.e. if and only if

\[
\alpha \tau \Pi \geq Z - k(\alpha)
\]

Let us denote \( \tau_\alpha \) the project quality such that the equality holds and study the variations of \( \tau_\alpha \) with \( \alpha \).

\[
\tau_\alpha = \frac{Z - k(\alpha)}{\alpha \Pi}
\]

\[
\frac{d\tau_\alpha}{d\alpha} = \frac{-\alpha k'(\alpha) - Z + k(\alpha)}{\alpha^2 \Pi}
\]

Typically, \( \tau_\alpha \) will first be increasing and then decreasing in \( \alpha \). This is due to

\[\text{footnote}{For instance, this could be due to imperfections on the capital market so that the large shareholder has to pay} \ (1 + \rho)v \ \text{to buy a share worth} \ v. \ \text{The takeover cost is then} \ k(\alpha) = \rho(1 - \alpha)v.\]
the combination of two conflicting effects. On the one hand as $\alpha$ increases takeover costs decrease and it becomes easier for the large shareholder to impose her authority. On the other hand, as $\alpha$ increases the large shareholder values security benefits increasingly more. This aligns her preferences with those of the manager.

Hence, the manager’s utility is

$$U_M = e[1 - F(\tau_\alpha)]b - c \frac{e^2}{2}$$

and the FOC yields

$$ce_\alpha = b[1 - F(\tau_\alpha)]$$

$$\frac{de_\alpha}{d\alpha} = -\frac{b}{c} \frac{d\tau_\alpha}{d\alpha} f(\tau_\alpha)$$

Typically, $e$ is initially decreasing and then increasing in $\alpha$ (within the range that allows him to retain control). This is the result of two antagonistic effects. On the one hand, as $\alpha$ increases the takeover threat becomes more forceful because the takeover costs are lower. On the other hand the large shareholder’s interest converge more with those of the manager.

The large shareholder gets the private gains whenever the manager does not find a project or when his project yields too low security benefits. Hence, the premium for the block is

$$p_\alpha = Z[1 - e + eF(\tau_\alpha)]$$

$$\frac{dp_\alpha}{d\alpha} = Z \frac{b}{c} [1 - F(\tau_\alpha)] f(\tau_\alpha) \frac{d\tau_\alpha}{d\alpha}$$

Typically, the premium for the block is first increasing and then decreasing in $\alpha$. The same factors which affect $e_\alpha$ also affect the premium although in the opposite direction. As $\alpha$ increases the takeover threat operates more easily due to the decreasing takeover costs. However, the large shareholder’s interests become more aligned.
This simple model does not intend to be a predictive theory of private gains. Nevertheless, it illustrates that the large shareholder’s private benefits depend on how easily or how much she can influence the management’s decisions. Furthermore, the feature that this influence increases with the size of the large shareholder’s block seems robust.

1.10 Comments

- The current view on the premium for a minority block is as follows. The degree of control an investor commands depends upon the strategic importance of her block in forming a controlling coalition. The larger her block is the larger is her bargaining power vis-à-vis other (large) shareholders in the division of private gains. 22.

In the last two sections, we propose two alternative (logical) relations between premium and block size. In contrast to the current approach which takes the total size of private gains to be shared as given, we obtain an endogenous control value of owning a minority block.

The first relation, obtained in section 1.8, relies on the possible conflict of interests between large and small shareholders. When a large shareholder has not full formal authority because she owns less than 50% of the votes, the larger her block is, the higher the degree of real authority she commands. For example, the board of directors is more likely to accept her recommendations if she has a larger stake in the firm, or the disclosure of information could be more credible when the large shareholder holds, say 30%, rather than 10% of the shares.

The second relation, obtained in section 1.9, relies on the conflict between the large shareholder and the manager, while the small shareholders are assumed to be passive. The large shareholder can threaten to withdraw by a takeover the manager’s real authority delegated by the current ownership structure. The larger the block is she owns the less costly it is replacing the manager. As a consequence, the larger the

large shareholder is, the more her preferences influence the decisions implemented by the manager.

- Reversing the perspective, our approach can provide some new insights into several related questions. Firms tend to fund most of their investment projects through retained earnings or debt. When and why do some firms resort to (voting) equity finance? In particular, what does trigger an Initial Public Offerings (IPO)? Our point is that a less concentrated ownership structure acts as a commitment by the large(r) shareholder(s) vis-à-vis outsiders to delegate some authority to the management and not to intervene constantly in order to pursue her private interests.

One possible story is the following. An entrepreneur developed his firm through debt. Thereby he kept close control over its operations, i.e. delegated very little to other managers. At some point, the cost of debt has become very high. The entrepreneur holds control rights, hence has formal authority, but a small fraction of return rights. He might then choose to issue equity in order to leave the management more freedom of action. At least, the management can be influenced by parties other than the entrepreneur. This in turn can relax the cost of debt.

Another conceivable story is that for exogenous reasons, outsiders tend not to trust insiders any more. Say, most of the firm shares are in the hands of a family. After the initial entrepreneur retired or when the environment changes drastically, it is no longer clear whether the family is competent in running the firm. Outsiders would then prefer the firm to switch from an entrepreneurial to a more managerial type (the firm has now many competent managers). 23

Going public can act as a commitment vis-à-vis the management not to be able to interfere too much. In the spirit of section 1.8, bringing outsiders inside the firm creates potential conflicts of interests. We suggest that they might turn out to be productive in terms of incentives of the management.

23 See the resistance of outsiders to Rupert Murdoch’s plan to issue stock with increased voting power to his own family before he retires. See The Economist, 23-29/10/1993.
1.11 The impact of a fiduciary duty

What is the difference between having a division entirely owned by a firm and having a majority block in a traded subsidiary? Although the allocation of formal authority is the same in these two ownership structures we believe that an important difference exists with regard to fiduciary duty. In a traded subsidiary fiduciary duties re-introduce the conflict of interests among shareholders by obliging the majority shareholder to act in the interests of all shareholders. In a non-traded division which is entirely owned by a firm this obligation of the controlling party is vacuous. We like to suggest that the form and the extent of the fiduciary duty affects the monitoring behaviour of the controlling party and thereby also the behavior of monitored party.

Given that the large shareholder has formal control and that there is a potential conflict of interests the question arises whether she should be unconstrained in exercising her controlling powers or rather subject to liability. In the latter case, the scope of these fiduciary duties needs to be defined. Within our framework one can think of two types of liabilities the large shareholder might be subjected to.

1. Negligence: She did not perform enough monitoring. Such a fiduciary duty deals with the contractible part of the monitoring activity.

2. Diversion of resources: She took the wrong decision even though she knew it was not maximizing security benefits. Such a fiduciary duty requires that the information the large shareholder had at the time she took the decision may be verified at least with some probability.

The two different types of fiduciary duty lead to very different conclusions. In particular, the first type of fiduciary duty leads obviously to higher levels of monitoring. Less obvious is the fact that the second type of fiduciary duty lead to lower levels of monitoring and thus to higher managerial initiative.
1.11.1 A simple model

To study the effects of fiduciary duties we use the model of section 1.8 and replace the minority shareholder with a majority shareholder. That is, the large shareholder commands now formal authority. Assume further that monitoring produces security benefits $T$ and that the large shareholder has an interest in keeping the incumbent manager if he is not too incompetent. Say, they have some sort of sweetheart deal which is worth $Z > \alpha p\Pi_2$ to the large shareholder. However, if the large shareholder finds evidence that the manager is incompetent and he conceals the information, there is a probability $\lambda$ that this information leaks out. In this case she is brought to court where she has to pay a penalty $P$.

For simplicity we assume in this section that $\alpha$ is given and we compare two different regimes.

No Fiduciary Duties

Given that the large shareholder is not subject to any fiduciary duty she is entirely free whether or not to make use of the acquired information. Furthermore she prefers to retain all but the very incompetent managers as $Z > \alpha p\Pi_2$. That is, only bad manager who produce a low signal and whose type the large shareholder does not learn are dismissed. Hence, the manager maximizes

$$U_M = b\{p + (1 - p)(e + (1 - e)E)\} - c\frac{e^2}{2}$$

and the FOC yields

$$ce = b(1 - p)(1 - E)$$

The large shareholder maximizes

$$V_L = \alpha ET + p[\alpha(\Pi_1 + \Pi_2) + Z] + (1 - p)e(Z + \alpha \Pi_1) + (1 - p)(1 - e)[EZ + (1 - E)\alpha p\Pi_2] - k\frac{E^2}{2}$$
and the FOC yields

\[ kE = \alpha T + (1 - p)(1 - \epsilon)(Z - \alpha p \Pi_2) \]

Despite the fact that there is a conflict of interests the manager's effort level is below the one in section 1.8. This is due to two facts. First, the small shareholders who would like to dismiss any manager with a low signal do not have the authority to do so. Second, the large shareholder has the authority but prefers to retain him, given she has learned his type.

**Fiduciary Duties**

As mentioned earlier, we analyze an extreme case of fiduciary duty in which the large shareholder never lies.

**Assumption 5** The fact that the large shareholder conscientiously took the wrong decision leaks out with probability \( \lambda \). In this case she has to pay a penalty equal to \( P \). We assume

\[ Z - \lambda P < \alpha p \Pi_2 \]

The high expected penalty the large shareholder has to pay for hiding information induces her to use all the information she has acquired. That is, whenever she learns that a manager is a bad type she dismisses him in order not to breach her fiduciary duties.

Hence, the manager's FOC is equal to

\[ ce = b(1 - p)(1 - E) \]

The large shareholder maximizes

\[ V_L = \alpha ET + p[\alpha(\Pi_1 + \Pi_2) + Z] + (1 - p)e[\alpha \Pi_1 + E\alpha \Pi_2 + (1 - E)Z] + (1 - p)(1 - \epsilon)\alpha p \Pi_2 - k \frac{E^2}{2} \]
and the FOC yields

\[ kE = \alpha T - (1 - p) [Z - \alpha \Pi_2] e \]

Given that the large shareholder derives private benefits only if the manager is retained we have the following result.

**Proposition 5** The monitoring effort is lower when there are fiduciary duties of the large shareholder than when there are no duties.

This result is surprising at first glance, but in accordance with the assumption that the large shareholder is heavily penalized when she conceals relevant information. Whenever she learns that the manager is a bad type she has to give up private benefits to avoid a possible penalty. Hence, monitoring is less valuable for the large shareholder and she prefers not to be involved too closely in the firm. Obviously this result depends on the nature of fiduciary duties we assume. If, for instance, the fiduciary duty is to ensure a minimum level of checking, the results are reversed.

Our approach offers an explanation of the trade-offs involved in publicly trading a subsidiary which is complementary to the existing theories. According to Holmström and Tirole (1993) public trading of a subsidiary creates additional information about a firm's performance which allows to design more efficient managerial contracts. The costs of market monitoring are that the initial buyers of the stock (liquidity traders) require a discount in order to be compensated for the losses they incur later on when trading with informed traders. Or, alternatively, there is a cost of registration and disclosure associated with taking the firm public. An alternative reason why publicly trading a subsidiary might be advantageous is based on the fiduciary duty a majority shareholder has towards minority shareholders. The obligation to act in the interests of all shareholders can be used as a commitment device to put the management under a tougher incentive scheme or to boost initiative. However, there are also costs associated with lower levels of monitoring, like retaining unqualified managers.
1.12 Extensions and future research

1.12.1 The rigidity of ownership concentration

We have argued that the optimal ownership concentration may differ according to the situation. In a dynamic framework, it should then vary across time, as the firm's environment, activities and reputation evolve. However, there are rigidities making such variations uneasy.

Grossman and Hart (1980) argue that the free-rider problem among small shareholders will prevent an outsider from capturing potential increase in security benefits he would implement. Shleifer and Vishny (1986) extend this argument by showing that a large shareholder partly internalizes the externality and thus make a profitable tender offer. Both these arguments carry over to the case of a mere attempt to concentrate the ownership, i.e. to buy or enlarge a block. As a result, a completely dispersed ownership is rather stable and a move towards a concentrated ownership is more likely when the initial ownership structure is already fairly concentrated. 24 This has several consequences.

On the one hand, the commitment to delegate some (real and possibly formal) authority via a partly dispersed ownership is fairly robust. It is costly to switch to a more concentrated ownership at a later stage. Hence, a dispersion of ownership (such as in equity carve out) is likely to last, and thus should induce the management and other employees to engage into firm specific investment that they value most when they have authority. 25 The robustness of the commitment may also be important vis-à-vis outside investors, such as debtholders (see section 1.10).

On the other hand, the optimal ownership structure accounts for this rigidity. Hence, with respect to the (static) optimal level, firms will tend to be over concen-

---

24 Note that the argument does not extend to the commitment to maintain a concentrated ownership and so the large shareholder's commitment to retain a large stake may be subject to a credibility problem.

25 Habib (1993) analyzes leveraged buyouts as a means to breach implicit contracts between the management and employees or business partners of the firm.
trated initially and then to be under concentrated. If there are some fixed costs of concentration (say, administrative costs of organising the purchase of a block) then one should observe over concentration of ownership following from a concentration phase.

1.12.2 Debt structure and design

In Burkart et al. (1994), we apply our ideas to the debt and design of debt. In particular, we start from two observations. First, banks are very often senior and secured creditors. Bondholders are usually junior creditors and are almost never secured. Second, the conventional wisdom\textsuperscript{26} is that banks are better monitors than the market.

These two observations are generally thought to be incompatible. If banks are superior monitors, they should be junior claimants and not secured in order to maximize their incentives to monitor. Being senior creditors gives the banks high incentives to liquidate firms in financial distress although continuation via refinancing might result in higher total expected returns. We propose a theory that reconciles the two observations above. Our idea is that banks can be excessively prone on restructuring the firm rather than liquidating it. This may well be optimal ex-post, but it reduces the incentive for the management to work hard ex-ante. This is an instance of the soft-budget constraint whose first formulation is due to Kornai (1980). To avoid the risk of the soft budget constraint, the bank’s payoff under liquidation needs to be increased. Security and seniority are two ways to achieve this goal.

1.13 Concluding Remarks

This paper provides a new perspective for the study of the optimal ownership structure of the firm. We stress a basic trade-off between a moral hazard and an adverse selection problem. The presence of a large shareholder provides through monitoring

\textsuperscript{26}See Hoshy, Kashyap and Scharfstein (1990) for an empirical investigation.
a higher quality of the management team's projects, but at the cost of an increasingly softer budget constraint. We then go on to analyze some applications of the main theory, trying to provide alternative explanations for the decision of going private/public, for the decision of opening a new division inside the firm vs. having a traded subsidiary, and for the privatization policy. Some important issues have been neglected. In particular, we believe that it is important to focus on the issues that arise when the state is the large shareholder and the firm is a public firm or a private regulated firm. Moreover, we ignore the role of the stock market as a monitoring device 27 in addition to the monitoring by the large shareholder. These extensions will (hopefully) be the object of future research.

Chapter 2

Debt Design, Liquidation Value, and Monitoring (joint with M. Burkart and D. Gromb)

2.1 Introduction

In this paper we study the design of contracts in a borrower-lender relationship. More precisely, we investigate how differences in the lenders' monitoring abilities and incentives interact with the liquidation value of the firm and with its division among different claimholders.

Two observations provide the motivation for our study. First, private debt and bank debt are often senior to public debt and secured\(^1\). In the sample analyzed by Asquith, Gertner and Scharfstein (1994), 51.5% of debt is publicly held and 48.5% is privately held. Private debt is held by banks (60.1%) or other financial institutions. Public debt is typically subordinate to private debt. Only a small fraction (10.5%) of private debt is centrally collateralized and public debt is generally denominated in dollars.

\(^1\)As Emmerick and White (1992) write, in the U.S. the primary distinction between private and public debt is that private placements are exempt from registration with the SEC, and thus from the costs of initial disclosure and continuous reporting requirements that go along with it. The rationale for this exemption is that institutional investors possess the necessary sophistication to invest directly - i.e. without the benefit of SEC-mandated information - in companies. As Carey, Prowse and Rea (1993) point out, the major investors in private placements are life insurance companies, which indeed invest in analyzing the companies to whom they lend.
of the public debt is secured while 55.5% of the private debt is secured. Similarly, Gorton and Kahn (1994) report that 70% of commercial loans are secured, the rest being tightly protected by covenants.²

The second observation is that, in the conventional wisdom, financial institutions like banks and insurance companies perform some monitoring of the borrowers that the market does not. Studying a sample of privately placed debt agreements, Kahan and Tuckman (1993) find that they require more monitoring and facilitate renegotiation compared to public bonds. Also, the ownership of private debt is often concentrated, a factor that enhances monitoring by mitigating free-rider problems. This feature of public debt has been confirmed by Hoshi, Kashyap and Scharfstein (1990) for financially distressed firms in Japan. They report that Japanese firms, which belong to a group (keiretsu) invest and sell more than non group firms in the years following financial distress.

The first observation seems to conflict with the logic of the second, as Gorton and Kahn (1994) notice:

The role of banks as ex post monitors suggests that banks should be junior claimants (and perhaps equity claimants) because their incentive to monitor would then be strongest [...] But, in fact, banks are typically senior, secured, claimants. It seems difficult to reconcile this feature of bank loans with the bank’s role as ex post monitor.

We propose a theory which reconcile the fact that private debt and bank debt are more protected in bankruptcy than public debt, even though the superior protection enjoyed by those creditors stifles their incentives to monitor.

We start from the consideration that information acquisition, i.e. monitoring, does not only come with benefits. That is, information enables the principal to renegotiate the initial contract, thereby ensuring an ex-post efficient decision. However, the potential for renegotiation reduces the disciplinary power of the liquidation threat,

²See also Gilson, John and Lang (1990).
as inefficient liquidation becomes less likely. Thus, enhancing ex-post efficiency may harm ex-ante efficiency. Consequently, it may be optimal not to facilitate and incentive monitoring too much. In particular, institutions may be designed so as to attain an optimal degree of monitoring. This point has been made recently in a few papers. For instance, Aghion and Tirole (1993) show that a principal can gain by committing not to monitor the agent’s activity. By remaining uninformed the principal commits not to overrule the agent. This in turn increases the agent’s incentives to exert effort.

An immediate consequence of the trade-off between ex-ante and ex-post efficiency associated with monitoring is that the amount of information, which strikes a balance between costs and benefits, depends upon the specific circumstances. That is, different situations may require different degrees of monitoring. Hence, there is scope for devices that induce different degrees of monitoring, contingent on the environment.

We focus on three such instruments: concentration and design of claims, and liquidation value. On the one hand, there are states where monitoring is unconditionally beneficial. In those states it is optimal to have a highly concentrated structure of claims. On the other hand, there are states in which liquidation is preferable since continuation, although optimal ex-post, destroys ex-ante incentives. In those states it is optimal to have a dispersed structure of claims. In general, this trade-off will lead to an optimal degree of concentration of claims. However, as a high liquidation value is a way of obtaining the benefits of monitoring without destroying ex-ante incentives, things can be improved. More precisely, the amount of monitoring depends on the difference between continuation and liquidation values. Increasing the liquidation value for the monitors allows to have more monitoring in the good states without distorting incentives in the bad states.

The paper is organized as follows: in the first part, we take concentration of claims as given and analyze the relationship between the project’s liquidation value and the lender’s monitoring ability. In particular, we assume that there is only one.

3See also Crémer (1993), Dewatripont and Maskin (1995) and Gertner et al. (1994).
creditor who demands a collateral for his loan. When the collateral available with
the entrepreneur's preferred project falls short of the amount of collateral required
by the creditor, the former is forced to distort the project choice. In order to get the
financing, the entrepreneur needs to choose a project with a lower expected return
but whose inputs contain sufficient inside collateral, i.e. assets which can be pledged.

We show that having a high ability to monitor might (optimally) require to distort
the project choice so as to reduce incentives to monitor. In other words, to counter
the adverse effects of monitoring, it might be necessary to bias the choice towards
projects with high liquidation value. This result is consistent with the observation
that banks seem to be much concerned with having collaterals for their loans but very
little concerned with the expected returns of projects.

In the second part we take the project choice as given and focus on a debt design
problem. In particular, we optimize both for the concentration of claims and for
the share that each claimholder should receive in liquidation. We conclude that
the monitoring claimholders should receive a larger share in liquidation than non-
monitoring claimholders. This resolves the apparent puzzle of private debt being
senior with respect to public debt.

The three papers which are closest to ours are Diamond (1993), Rajan and Win-
ton (1994) and Dewatripont and Tirole (1994). Diamond (1993) argues that it is
optimal to structure short-term debt to be senior to long-term debt. This arrange-
ment precludes concessions on the part of short-term lenders which in turn leads
to efficient ex-post liquidation policies. Since banks are typically short-term credi-
tors, he also reaches the conclusion that banks should be senior claimants. Notice,
however, that the two explanations have little in common: we derive the result that
monitors should be senior, irrespective of the horizon of the financing relationship (in
our model all claims have the same maturity). Since the conventional wisdom sees
banks as monitors, we can interpret our results as saying that banks should be senior
claimants.

The two explanations have different implications in the case of private non-bank
debt. Emmerick and White (1992) provide evidence that this type of relationship is
of medium-long term, but still lenders are very protected in case of default. This observation is compatible with our explanation, but not with the one given by Diamond (1993).

Rajan and Winton (1994) start from the same consideration that financial institutions monitor their borrowers, and that the actual level of monitoring depends positively on the size of their claims. However, their view on the function of monitoring differs from ours. According to Rajan and Winton monitoring prevents inefficient continuation. Consequently, loan contracts must be structured to enhance the lender’s incentives to monitor. Covenants and collateral achieve this goal by making the effective priority of the loan contingent on monitoring. Our theory associates monitoring with excessive continuation rather than more frequent termination. Collateral is a means of obtaining the benefits of monitoring without incurring the problem of excessive softness in financial distress.

In Dewatripont and Tirole (1994) control is allocated to bondholders after a poor first-period performance. Bondholders have (by assumption) higher incentives to liquidate the firm than the equity-holders. This potential control shift strengthens the management incentives to exert high effort. Our model shares with theirs the idea that high incentives to liquidate ex-post can be beneficial ex-ante.

2.2 The model

In this section we present the basic model. We will be using variations on it in the rest of the paper. We borrow from many, mainly from Crémer (1993) and Dewatripont and Maskin (1995).

At date -1, a penniless entrepreneur has a project that needs inputs, funding, and management.

Management: The entrepreneur hires a manager to run the project. The manager is indispensable and cannot be replaced.

Funding: The entrepreneur borrows the set-up cost I from a single creditor, hence-
forth the bank. See section 5 for a study of multiple creditors. The two parties sign a debt contract, with face value $D$. The capital market is competitive, and hence the bank earns zero profit in equilibrium. For simplicity, we assume that the interest rate is zero.

**Inputs:** The entrepreneur builds the productive capacity from two types of assets, say physical capital and human capital. He invests $L$ in the former asset and $I - L$ in the latter. For simplicity, we assume that the two assets have equal prices. The input mix chosen by the entrepreneur is verifiable.

**At date 0,** the manager chooses a level of effort, $e \in \{\underline{e}, \bar{e}\}$, with $\underline{e} < \bar{e}$. The additional disutility of exerting the high effort $\bar{e}$ is $c > 0$. The effort level is not verifiable.

**At date 1,** the project can be either quick, paying $R_q = f(L)$ where $f(L)$ is a concave function, or slow, paying 0, with probability $e$ and $(1 - e)$ respectively. In the former case the game ends 4. In the latter case, the firm is in financial distress, and the creditor has the right to decide whether to continue or to liquidate.

**At date 2,** returns accrue depending on the decision to continue or liquidate the firm.

**Liquidation after default:** The two assets differ in their liquidation value. Physical capital, like machinery or inventories, can be posted as collateral and has a positive liquidation value, while human capital cannot be pledged. In addition, we assume 5 that there is a deadweight loss of $(1 - \beta)$, with $\beta \geq 0$, involved in the liquidation of the physical assets. Hence, the liquidation value for the creditor is $\beta L$.

Given that $f(L)$ is the assets' most productive use, and that liquidation is always feasible, we make the following assumption for consistency.

**Assumption 6** For all $L \in [0, I]$, $f(L) \geq \beta L$.

**Continuation after default:** If continuation is chosen, the liquidation value falls to

---

4 We can think of $R_q$ as the sum of cash-flow and liquidation value.

5 As in Bester (1985).
zero. Continuation yields a safe return of $R_0$. However, the bank can improve the continuation prospects by engaging in monitoring. At a cost $\rho E$, with $E \in [0, 1]$, an alternative continuation plan is found with probability $\sqrt{E}$, which generates a return $R_0 + R$, with $R > 0$.

The manager earns private benefits $B$ when the project is quick or when the slow project is continued. If the creditor decides to terminate the slow project, he gets zero private benefits. For simplicity, we rule out monetary incentive schemes.

2.3 Monitoring and the soft budget constraint

In this section, we explore the basic trade-off associated with monitoring. Given that the manager is indispensible for the firm to operate, the liquidation threat is necessary to discipline him. Hence, the creditor would like to commit to a hard budget constraint such that default triggers liquidation. However, when contracts are not binding, the creditor may have an incentive to renege on this arrangement and continue rather than liquidate the firm. Whether it is willing to renegotiate depends on how returns under continuation and liquidation compare. Since the continuation value is decreasing in monitoring costs, a high ability (or incentive) to monitor may yield a soft budget constraint. To keep the analysis of this basic trade-off simple, we take the amount of $L$ as given in this section. Additionally, we will always implicitly assume that $f(L)$ is high enough to ensure that the project is feasible, i.e. can attract the required funds.

We now compute the expected value of continuation after default.

Lemma 2 After default, the continuation value is

- $C(\rho) = R_0 + R - \rho$ if $\rho \leq R/2$;

---

6This is a typical instance of soft budget constraint. As Maskin (1994) writes, “the soft budget constraint syndrome [...] pertains whenever a funding source - e.g. a bank or government - finds it impossible to keep an enterprise to a fixed budget. That is, the enterprise can extract ex-post a bigger subsidy or loan than would have been considered efficient ex-ante.”
• \( C(\rho) = R_0 + R^2 / 4\rho \) if \( \rho \geq R / 2 \).

**Proof** When monitoring, the creditor maximizes \( R_0 + \sqrt{E} R - \rho E \). The FOC yields \( \sqrt{E} = \frac{R}{2\rho} \) for \( \rho \geq \frac{R}{2} \), and \( E = 1 \) otherwise.

Absent monetary incentives, only the liquidation threat can discipline the manager to exert the high effort. Because the manager is indispensable for the firm to operate, the firing threat is equivalent to the liquidation threat.

**Lemma 3** The manager exerts \( \bar{e} \) if and only if the firm is liquidated after default with a probability \( \mu \) greater or equal to \( \frac{c}{(\bar{e} - \underline{e})B} \).

**Proof** Since effort is non-contractible, the probability \( \mu \) of liquidation after default is independent of \( e \). The incentive compatibility-constraint for \( \bar{e} \) is

\[
\bar{e}B + (1 - \bar{e})(1 - \mu)B - c \geq \underline{e}B + (1 - \underline{e})(1 - \mu)B
\]

Obviously, if \( \frac{c}{(\bar{e} - \underline{e})B} > 1 \), there is no liquidation threat which could affect the manager's effort choice. In what follows, we restrict the analysis to cases where the liquidation threat, if credible, induces the manager to work hard.

**Assumption 7** \( \frac{c}{(\bar{e} - \underline{e})B} < 1 \).

By Lemma 2, as \( \rho \) goes to infinity, the continuation value tends to its minimum, \( R_0 \). If \( \beta L < R_0 \) the continuation value is greater than the liquidation value for all values of \( \rho \), and the firm is never liquidated after default. This soft budget constraint makes it impossible to implement the high effort. When \( \rho = \infty \) the face value of debt required by the bank to break even is

\[
\rho^* = \frac{1}{\underline{e}} [I - (1 - \underline{e})R_0]
\]

and the entrepreneur's corresponding expected return equals

\[
\underline{e}[f(L) - D] = \underline{e}f(L) + (1 - \underline{e})R_0 - I
\]
Conversely, as $\rho$ goes to zero, the continuation payoff tends to its maximum, $R_0 + R$. If $\beta L > R_0 + R$ liquidation dominates continuation for all values of $\rho$, and the firm is always liquidated after default. The hard budget constraint forces the manager to exert a high effort.

**Proposition 6** Assume $R_0 + R > \beta L > R_0$. The expected payoff of the entrepreneur is not weakly decreasing in the monitoring cost.

**Proof** As $\rho$ increases from 0 to infinity, the continuation value $C(\rho)$ decreases from $R_0 + R - \rho$ to $R_0$. Denote by $\hat{\rho}$ the monitoring cost defined by $C(\hat{\rho}) = \beta L$. For $\rho < \hat{\rho}$, liquidation is not credible, and the manager shirks. The entrepreneur’s payoff is

$$\epsilon f(L) + (1 - \epsilon)C(\rho) - I$$

For $\rho \geq \hat{\rho}$, liquidation dominates continuation and the hard budget constraint induces the manager to exert the high effort. The entrepreneur’s payoff is

$$\bar{\epsilon} f(L) + (1 - \bar{\epsilon})\beta L - I$$

At $\hat{\rho}$ the entrepreneur’s payoff jumps upwards by magnitude $(\bar{\epsilon} - \epsilon)[f(L) - \beta L]$. □

An increase in $\rho$ reduces the creditor’s continuation payoff. Since this increases the amount of debt $D$, the entrepreneur’s profit also decreases initially in the monitoring costs. However, when $\rho$ reaches the threshold level where $C(\rho) = \beta L$, continuation and liquidation yield the same return and the manager switches to the high effort level. Hence, the entrepreneur’s profit jumps upward. As $\rho$ increases further his profit does not depend on $\rho$ anymore, since the creditor always prefers to liquidate.

**Proposition 7** The entrepreneur’s payoff is higher for $\rho = +\infty$ than for $\rho = 0$ if and only if $(1 - \epsilon)[f(L) - (R_0 + R)] > (1 - \bar{\epsilon})[f(L) - \beta L]$.

**Proof** For $\rho = \hat{\rho}$, the entrepreneur’s payoff is $\bar{\epsilon} f(L) + (1 - \bar{\epsilon})\beta L - I$. For $\rho = 0$, the entrepreneur’s payoff is $\epsilon f(L) + (1 - \epsilon)(R_0 + R) - I$. □
For $\rho \geq \hat{\rho}$ the creditor prefers liquidation to continuation. The entrepreneur's profit is independent of $\rho$ and equal to $\bar{\epsilon} f(L) + (1 - \bar{\epsilon}) \beta L$. Thus, if the condition above is met, the entrepreneur is better off when the creditor is inefficient at monitoring, i.e. $\rho \geq \hat{\rho}$.

The basic message of this section is simple: monitoring does not only comes with benefits, but also with indirect costs. In particular, if we see monitoring as useful in situations of financial distress, excessive monitoring can lead to low effort on the part of the management team. The following implication illustrates the message of this section.

**Corollary 2** The entrepreneur may be best off with a positive monitoring cost.

It is worth pointing out that this view of monitoring is not the only possible one. In particular, an alternative would be to consider monitoring as the ability to influence the manager's choices so that he acts in the best interests of the shareholders. This view of monitoring would change our results. Our formalization of the monitoring process captures the feature that the firm's stakeholders intervene periodically or in crucial situations, like financial distress, and try to assess what is the best option for the firm.

### 2.4 Monitoring and Liquidation Value

This section explores the trade-off associated with monitoring further and argues that a high liquidation value may be necessary for liquidation to be credible. Hence, it may be best to have a high level of $L$, even at the cost of departing from the optimal input mix. That is, for a given monitoring ability, it might be optimal to distort the project choice towards one with lower expected return, but with higher liquidation value, in order to reduce the moral hazard problem. The previous section illustrates that low monitoring costs may soften the budget constraint. As a consequence, our model predicts that the lower the monitoring costs, the more assets that can be pledged are required to strengthen the budget constraint. $L$ is the key choice-variable in
the trade-off between strengthening the budget constraint and distorting the optimal input mix. Hence, we relax the assumption of the previous section where $L$ has been treated as exogeneously given. We will refer to $L$ as collateral. Notice, moreover, that in our model collateral is inside collateral, i.e. the pledged assets are firm assets, rather than the entrepreneur’s personal wealth.

For concreteness we use a specific form for the function $f(L)$. Our results hold for any concave function.

**Assumption 8** $f(L) = I + \lambda L(I - L)$, with $\lambda > 0$.

Let us first derive the optimal “undistorted arrangement” the entrepreneur would choose, absent any moral hazard problem. For a given input mix and effort $\bar{e}$, the entrepreneur’s profit is

$$\bar{e}f(L) + (1 - \bar{e})\max\{\beta L; C(\rho)\}.$$ 

Hence, conditional on $e = \bar{e}$ the optimal unconstrained amount of collateral $L$ is either

$$\arg\max\{f(L)\} = \frac{I}{2} \text{ for } C(\rho) > \beta L$$

or

$$\bar{L} = \arg\max\{\bar{e}f(L) + (1 - \bar{e})\beta L\} = \frac{I}{2} + \frac{(1 - \bar{e})\beta}{2\lambda \bar{e}} \text{ for } C(\rho) \leq \beta L.$$ 

To implement $\bar{e}$, the liquidation threat has to be credible. Hence, it must be the case that $\beta L \geq C(\rho)$. The idea we want to capture is that the cost of implementing the high effort is a distorsion of the input mix. Remember that $C(\rho)$ decreases from $R_0 + R$ to $R_0$ as $\rho$ increases from 0 to infinity. If $\beta \bar{L} > R_0 + R$ then $L = \bar{L}$ for all $\rho$, and the implementation of $\bar{e}$ does not involve any distorsion. If $R_0 + R > \beta \bar{L} > R_0$, the implementation of $\bar{e}$ implies $L > \bar{L}$ for low values of $\rho$ and $L = \bar{L}$ for high values of $\rho$. Only if $R_0 \geq \beta \bar{L}$ the input mix has to be distorsed with respect to $\bar{L}$ for all $\rho$. That is, it is necessary to set $L = C(\rho)/\beta > \bar{L}$. In what follows we restrict ourselves to this

---

\footnote{Note that Assumption 6 is satisfied for all $\beta \in [0, 1]$ as for all $L \in [0, I], f(L) \geq I \geq \beta L$.}
last case. The assumption \( R_0 \geq \beta \bar{L} \) for all \( \beta \in [0, 1] \) allows us to derive comparative-statics results with respect to \( \rho \) and \( \beta \).

**Assumption 9** \( R_0 \geq \beta \bar{L} \)

**Lemma 4** If \( C(\rho) > \beta I \), it is impossible to implement \( \bar{\varepsilon} \). Otherwise, the optimal amount of collateral needed to implement \( \bar{\varepsilon} \) is

\[
L = \frac{C(\rho)}{\beta} > \bar{L}
\]

**Proof** \( \beta L \geq C(\rho) \) is necessary for the liquidation threat to be credible. Since \( \bar{\varepsilon}f(L) + (1 - \bar{\varepsilon})\beta L \) is concave, the constraint is binding.

When the creditor is relatively efficient in monitoring, continuation dominates liquidation after default, and the manager shirks. However, by distorting the input mix in favour of the collateralizable asset, i.e. by increasing \( L \), liquidation becomes more attractive to the creditor. When the liquidation value \( \beta L \) reaches the continuation value, the manager reacts by choosing the high effort level.

**Lemma 5** The optimal amount of collateral needed to implement \( \varepsilon \) is \( L = \frac{I}{2} \).

**Proof** Since \( R_0 \geq \beta \bar{L} \) (by Assumption 9) and \( \bar{L} > I/2 \), we have for all \( \rho \), \( C(\rho) > \beta I/2 \). Hence, setting \( L = I/2 \) triggers the soft budget constraint and implements \( \varepsilon \).

When the creditor is an efficient monitor the implementation of \( \varepsilon \) does not create any problems. Continuation ensures that the manager prefers the low effort level, and there is no reason to deviate from the optimal input mix \( I/2 \).

From Lemma 4 and 5 follows that the entrepreneur's profit when the firm is liquidated after default, \( \Pi(\bar{\varepsilon}) \), and his profit when the firm is continued, \( \Pi(\varepsilon) \), have the following form. (Note that \( \beta L = C(\rho) \).

\[
\Pi(\bar{\varepsilon}) = \bar{\varepsilon}f\left(\frac{C(\rho)}{\beta}\right) + (1 - \bar{\varepsilon})C(\rho) - I
\]
\[ \Pi(\varepsilon) = \varepsilon f\left(\frac{I}{2}\right) + (1 - \varepsilon)C(\rho) - I. \]

The following two Lemmata allow us to derive the section’s main Proposition.

**Lemma 6** As \( \rho \) increases \( \Pi(\overline{\varepsilon}) \) increases and \( \Pi(\varepsilon) \) decreases.

**Proof** \( \overline{\varepsilon} f(L) + (1 - \overline{\varepsilon})\beta L - I \) is a concave function of \( L \) with its maximum at \( \overline{L} \). As \( L = C(\rho)/\beta > \overline{L} \), \( \Pi(\overline{\varepsilon}) \) increases as \( C(\rho)/\beta \) decreases, i.e. moves towards the optimum \( \overline{L} \). \( \Pi(\varepsilon) \) is strictly decreasing in \( C(\rho) \). \( \Box \)

**Lemma 7** As \( \beta \) increases, \( \Pi(\overline{\varepsilon}) \) increases and \( \Pi(\varepsilon) \) remains constant.

**Proof** \( f(L) \) is a concave function with its maximum at \( I/2 \). As \( L = C(\rho)/\beta > \overline{L} > I/2 \), \( \Pi(\overline{\varepsilon}) \) increases as \( C(\rho)/\beta \) decreases, i.e. moves towards the optimum \( I/2 \). \( \Box \)

**Proposition 8** (a) Suppose \( \overline{\varepsilon} f(R_0/\beta) - (\overline{\varepsilon} - \varepsilon)R_0 > \varepsilon f(I/2) > \overline{\varepsilon} f(R_0/\beta) - (\overline{\varepsilon} - \varepsilon)(R_0 + R) \).

There exists a unique \( \hat{\rho} \) such that, for \( \rho \leq \hat{\rho} \) it is optimal to set \( e = \varepsilon \). For \( \rho > \hat{\rho} \), it is optimal to set \( e = \overline{e} \).

(b) Suppose \( \overline{\varepsilon} f[C(\rho)] - (\overline{\varepsilon} - \varepsilon)C(\rho) > \varepsilon f(I/2) \).

There exists a unique \( \hat{\beta} \) such that, for \( \beta \leq \hat{\beta} \) it is optimal to set \( e = \varepsilon \). For \( \beta > \hat{\beta} \), it is optimal to set \( e = \overline{e} \).

**Proof** By the previous Lemmata, \( g(\rho, \beta) = \Pi(\overline{\varepsilon}) - \Pi(\varepsilon) \) is strictly increasing in both arguments. (a) The first inequalities express \( g(0, \beta) < 0 \) and \( \lim_{\rho \to +\infty} g(\rho, \beta) > 0 \). (b) The second inequality expresses \( \lim_{\beta \to 1} g(\rho, \beta) > 0 \). For \( \beta \) small enough, implementing \( \overline{\varepsilon} \) is either inferior to implementing \( \varepsilon \) or infeasible. \( \Box \)

Notice that conditions (a) and (b) do not hold simultaneously for \( \rho \) close to 0. They should be interpreted as separate conditions.

The result proved above implies a non-monotonic relationship between the efficiency of banks as monitors and the amount of collateral needed in equilibrium. In
particular, it is not generally true that more inefficient monitors will require more collateral. This result has an interesting application as regards the difference between the amount of collateral required by banks and that required by dispersed bondholders. In the model above disperse bondholders can be represented by a very high value of $\rho$. Suppose, for instance, that monitoring involves a fixed cost. Only the creditors who have a sufficiently high stake in the project find it profitable to acquire the information. Otherwise, they will simply decide not to monitor the project. The latter is formally equivalent to setting $\rho = +\infty$. However, as we will show later in the paper, this non-convexity argument is not crucial.

In this section we have established a direct relationship between the monitoring ability of creditors and the liquidation value of the project. Our result on the distortion of the optimal input mix may be thought as unrealistic. However, it can be reinterpreted as stating that banks refuse to finance projects with low liquidation value, even if their return prospects are good. This interpretation is in accordance with the often heard complaint by entrepreneurs that banks are interested in collateral and not in projects. The usual explanation views this phenomenon as the result of bank inefficiency and conservative behavior. In contrast, our theory explains why such behavior by the banks may be optimal.

This model differentiates itself from numerous other models on collateral by assuming that the pledged assets are owned by the firm rather than privately by the entrepreneur. The basic role of outside collateral is to increase the punishment for default. When a business fails, the entrepreneur loses (part of) his private wealth, in addition to forgoing any return. This threat can be used to enforce the selection of low-risk projects, e.g. Bester (1985), or to reduce the debtor’s incentive to declare strategic default (Bester (1994)). The purpose of collateral in the present model is to overcome a moral hazard problem. Collateralization, in the sense of choosing a project with a high liquidation value $\beta L$, makes the liquidation threat credible, and thereby prevents the manager from shirking. Outside collateral cannot perform this function. Suppose that absent any collateral the creditor chooses continuation after default.
The introduction of outside collateral does not affect the difference between continuation and liquidation values, and the creditor still prefers continuation to liquidation. In the event of default, the creditor simply seizes the outside asset and continues the firm. Hence, outside collateral does not affect the manager’s effort choice. Its only effect is to reduce the face value of debt. Due to the dead-weight loss \((1 - \beta)\), associated with the seizure of the pledged asset, a debt contract with outside collateral is dominated by unsecured debt.

This conclusion does not depend upon the manager’s limited liability. The creditor’s continuation decision remains unaffected, irrespective of whether the outside collateral is the owner’s or the manager’s personal wealth. However, in a setting where the party who makes the effort choice pledges personal wealth, outside collateral can resolve the moral hazard problem, albeit in a different manner. Consider a firm with a owner-manager. Assume that the owner-manager secures the debt by additional outside assets, instead of choosing a project with higher liquidation value, but lower expected return. Assume further that in the absence of the additional outside collateral the creditor opts for continuation after default. As before, the outside collateral does not change the creditor’s continuation decision, but only gives him an additional return in the event of default. Hence, outside collateral has no effect on the owner-manager’s incentives through the creditor’s decision to liquidate or continue. However, there is a direct effect. The risk of losing his pledged wealth provides the owner-manager with strong incentives to exert high effort.

The result that outside collateral fails to mitigate the moral hazard problem, when the agent enjoys limited liability, seems to be consistent with the syllized facts on the relative importance of inside and outside collateral. Outside collateral is more common in small business loans, while larger firms tend to secure their debt with corporate assets. Small firms are mostly run by their owners. The threat of losing some personal wealth in the event of default is an effective disciplining device for the entrepreneur. In contrast, larger companies are usually run by managers who enjoy limited liability. Collaterals have a disciplinary effect on the managers only if they affect the likelihood that the firm is liquidated. Hence, the pledged assets need to
be corporate assets, so that their seizure results in the firm’s liquidation. Finally, outside collateral seems also conceptually out of place for large companies. Imposing outside collateral either on the owners or the employees conflicts with the very idea of limited liability, a prominent contractual feature of large companies.

2.4.1 A reinterpretation: management entrenchment

As firing the manager is equivalent to liquidation in our model, it can be reinterpreted as a management entrenchment model. Shleifer and Vishny (1989)⁸ argue that a manager has incentives to invest the firm’s resources in assets whose value is higher under him than under the best alternative manager, even when such investments are not value-maximizing. As a result of such entrenching investments, replacing the manager is costly and he can extract a higher compensation from the shareholders. Shleifer and Vishny’s entrenchment result depends crucially upon the assumption that the information available to the monitoring board changes over time. On the one hand, the board is insufficiently informed at the time when the manager-specific investment is made and fails to intervene. On the other hand, once the investment is made, the board realizes that the value of the asset in place is higher under the incumbent’s control than under any alternative management team, and it retains the current management.

In our model, as in Shleifer and Vishny (1989), the asset composition determines the likelihood of retaining/firing the incumbent manager. However, in our model, it is the principal, and not the manager, who decides how to allocate the funds between the two assets. We show that, in the presence of managerial moral hazard, even the value-maximizing project can lead to management entrenchment ex-post, given that the creditor is an efficient monitor. So the entrepreneur, anticipating the problem, may choose an input-mix with a high liquidation value, i.e. higher redeployability of assets, in order to reduce his dependence on management.

If, instead, the manager could choose the input-mix, he would deviate in exactly

---

⁸A very interesting paper on the relationship between management entrenchment and bankruptcy procedures is Bebchuk and Picker (1994).
the opposite direction, i.e. ensure that the creditors always choose to continue.

In our model a high liquidation value is a way of reducing management entrenchment. Obviously liquidation is only one possible instrument of fighting management entrenchment (and an extreme one). But the spirit of the model is unchanged if we consider management replacement instead of liquidation. The message of the model would then be that it might be optimal to deviate from the project with the highest expected return if that project induces a management entrenchment problem, e.g. because there is a lot of specific knowledge accumulated in running it.

Moreover by stressing the role of the monitoring ability in determining the importance of the management entrenchment problem, our model goes one step further than Shleifer and Vishny (1989). The ideal situation in both models is when the monitor is unable to perceive ex-post the extra value under the current management. When the commitment not to be informed is impossible (or too costly) other instruments come into play to reduce managerial entrenchment.

2.4.2 Recessions and booms

It is a commonly accepted stylized fact that banks require more collateral in recessions than in booms. Our model is compatible with this observation. There are two ways to incorporate the difference between a recession and a boom in our model.

First, one can postulate that collateral is worth more during booms than during recessions i.e. that booms are associated with higher $\beta$ values than recessions. This interpretation is natural if we think of collateral as buildings. The demand for housing is generally higher during periods of high economic activity. When $\beta$ decreases initially the collateral required increases. This is in accordance with the stylized facts. There is, however, a further effect: when $\beta$ is very low, continuation dominates liquidation, the manager shirks, and no collateral is required.

Rather than interpreting the region of very low $\beta$ values as a deep recession, one may alternatively view it as a situation where the allocation of property rights and the legal system are very inefficient. This applies, for instance, to Eastern Europe during its current transition to a market economy.
A poorly defined property rights system makes the termination threat not credible, as it induces excessive softness towards management. Hence, our theory implies that the transition from centralized to decentralized economies in Eastern Europe will not be able to relieve the soft budget constraint problem unless it is matched with a better property rights system.

Another way of modelling the difference between recessions and booms is to assume that \((\bar{c} - \epsilon)^{reces} > (\bar{c} - \epsilon)^{booms}\). The condition states that during booms projects can succeed even when they are not carefully planned and executed, while during recessions bad projects are doomed to failure. In other words, the managerial moral hazard problem is exacerbated during recessions. In a recession, \(\Pi(\bar{c})\) shifts upwards compared to \(\Pi(\bar{c})\) in a boom. This implies that \(\hat{p}^{reces} < \hat{p}^{boom}\), and we have the following result:

A) for \(\hat{p} \leq \hat{p}^{reces}\), \(C = \frac{1}{2}\) in both recessions and booms

B) for \(\hat{p} \geq \hat{p}^{boom}\), \(C = \frac{R^2}{\hat{p}}\) in both recessions and booms

C) for \(\hat{p}^{reces} \leq \hat{p} \leq \hat{p}^{boom}\), \(C^{reces} = \frac{R^2}{\hat{p}}\) while \(C^{boom} = \frac{1}{2}\).

In both formulations, the amount of collateral required in recessions is higher than in booms, showing the consistency\(^9\) of our model to the stylized facts.

### 2.5 Multiple creditors

So far we have assumed that the firm had only one creditor. Introducing multiple creditors allows us to model another instrument to reduce the soft budget constraint problem created by excessive monitoring: the concentration of claims. When several creditors share the returns, monitoring is a public good. Each creditor tends to free-ride on monitoring by others. As a result, multiple creditors, as a group, monitor less than a single creditor who holds the sum of their claims. But this ex-post inefficiency can turn out to be productive because it gives stronger ex-ante incentives to the manager. The idea of this section is similar to the one in Burkart et al. (1994). We

\(^9\)It is worth mentioning that we have many parameters in our model, so that there is not a unique way of modelling recessions and booms. This makes our prediction not very sharp.
modify our previous model as follows.

**At date -1**, the entrepreneur can borrow from up to \( n \) creditors. We assume that only one type of claim is issued. This assumption is relaxed in the next section. Let \( \alpha_i \in [0, 1] \) denote the fraction of all outstanding claims held by creditor \( i \in \{1, \ldots, n\} \). For simplicity, assume that \( \alpha_1 > \alpha_2 > \ldots > \alpha_n \).

Each creditor \( i \) can exert a monitoring effort \( E_i \in [0, 1] \) at cost \( p E_i \). Let \( E = \min \{1, \sum_{i=1}^{n} E_i\} \) denote the creditors' joint monitoring effort. As before the creditors find a way to improve the continuation value from \( R_0 \) to \( R_0 + R \) with probability \( \sqrt{E} \). The creditors monitor simultaneously to the manager's effort choice without observing the latter's action. In case of default, the creditors can choose between liquidation and continuation. The former has a value \( \beta L \). The continuation value is either \( R_0 + R \) when the creditors found the improvement or \( R_0 \) if they were unsuccessful in their search.

In contrast to the previous sections, monitoring takes place at date 0, prior to the first period realization as well as prior to the decision to continue or liquidate. We choose monitoring ex-ante for several reasons. First, Lummer and McConnell (1989) provide evidence that a monitoring creditor acquires information about the firm during the relationship and not (only) when the firm is in financial troubles. In their study, Lummer and McConnell analyze the effect of bank loan announcements on the borrowers' stock price. They find significant stock price responses to revisions of bank loans, while new bank loans have no significant announcement effect. Their findings support the view that banks transmit firm-specific information which they had access to during their relationship with the borrowers. Thus, the information acquired during "normal" times is used when life gets tough. Second, the main virtue of the previous ex-post monitoring framework is to facilitate computing the cost of collateral, as it is independent of all efforts. Collateral is no longer our main concern. Third and more importantly, ex-post monitoring with several creditors raises the issue of the free-rider problem in concentrating claims ex-post (Grossman and Hart 1980). As our results rely on another instance of free-riding, this may confuse the reader. Ex-ante monitoring disentangle the two issues, and allows us to differentiate
our contribution more clearly from others based on Grossman and Hart’s free-rider problem. This issue is discussed in more detail later in the paper.

To simplify the exposition, we assume that monitoring is too costly to ensure that the improvement is found with probability 1.

**Assumption 10** \( \rho \geq \frac{(1-e)(R_0+R-\beta L)}{2} \)

To prevent that continuation is chosen with probability 1 despite the high monitoring cost we need to impose an additional restriction.

**Assumption 11** \( R_0 + R > \beta L > R_0 \)

We start by characterizing the equilibrium in the monitoring stage.

**Lemma 8** For a given \( e \), there exists a unique equilibrium among creditors. Creditor 1 is the only active monitor, i.e. \( \sqrt{E_1} = \alpha_1(1 - e)/(R_0 + R - \beta L)/2\rho \) and for all \( i \neq 1 \), \( E_i = 0 \).

**Proof** Given \( e \), creditor’s \( i \) problem is

\[
\max_{E_i} (1-e)\alpha_i[(R_0 + R) \sqrt{E_i + \sum_{j \neq i} E_j + \beta L(1 - \sqrt{E_i + \sum_{j \neq i} E_j})}] - \rho E_i
\]

subject to the constraint \( 0 \leq E_i \leq 1 \). Differentiating with respect to \( E_i \) yields

\[
(1 - e)\alpha_i(R_0 + R - \beta L) - 2\rho \sqrt{E}
\]

Since \( E < 1 \) by Assumption 10, in equilibrium

\[
(1 - e)\alpha_i(R_0 + R - \beta L) - 2\rho \sqrt{E} \leq 0.
\]

Hence, for all \( j \neq 1 \) \( (1 - e)\alpha_j(R_0 + R - \beta L) - 2\rho \sqrt{E} < 0 \), and \( E_j = 0 \). The only equilibrium is such that

\[
\sqrt{E_1} = \frac{(1 - e)\alpha_1(R_0 + R - \beta L)}{2\rho}
\]
and all the other creditors do not monitor.

The Lemma above illustrates the main effect of this section: creditors free-ride on each other in the provision of the public good. In particular, only the creditor with the biggest stake will be an active monitor. All other creditors abstain from monitoring because the marginal costs are the same for all creditors, while the marginal returns are increasing in the size of the creditor's stake. Hence, in equilibrium, the first-order condition can only be satisfied for the largest creditor.

**Corollary 3** For a given \( \varepsilon \), the continuation probability is increasing in creditor 1's stake \( \alpha_1 \).

**Proof** \( \sqrt{E} = \sqrt{E_1} = (1 - \varepsilon)\alpha_1 \frac{R_0 + R - \beta L}{2p} \) which is increasing in \( \alpha_1 \).

By Assumption 11, the firm is continued whenever the creditors find the improvement \( R \). As only creditor 1 monitors actively, and his monitoring effort is increasing in his stake \( \alpha_1 \), so is the continuation probability. We can now solve for the manager's optimal effort choice as a function of creditor 1's stake.

First define

\[
\alpha = \min \left\{ \frac{2p}{(1 - \varepsilon)(R_0 + R - \beta L)} \left(1 - \frac{c}{(\varepsilon - \varepsilon)B}\right), 1 \right\}
\]

\[
\bar{\alpha} = \min \left\{ \frac{2p}{(1 - \varepsilon)(R_0 + R - \beta L)} \left(1 - \frac{c}{(\bar{\varepsilon} - \varepsilon)B}\right), 1 \right\}
\]

Since \( \varepsilon > \bar{\varepsilon} \) we have \( \alpha \leq \bar{\alpha} \).

**Lemma 9** For \( \alpha_1 \in [0, \bar{\alpha}] \), there is a unique Nash equilibrium and \( \varepsilon \) is exerted.

If \( \alpha_1 \in [\alpha, \bar{\alpha}] \), there are several Nash equilibria. In particular, there are two pure strategy equilibria. In one of them, \( \varepsilon \) is exerted, while \( \bar{\varepsilon} \) is exerted in the other.

For \( \alpha_1 \in [\bar{\alpha}, 1] \) there is a unique Nash equilibrium and \( \varepsilon \) is exerted.

**Proof** By Lemma 3 effort \( \bar{\varepsilon} \) is chosen iff \( \sqrt{E_1} \leq 1 - \frac{c}{(\varepsilon - \bar{\varepsilon})B} \). Suppose \( e = \bar{\varepsilon} \). Creditor 1 sets \( \sqrt{E_1} = (1 - \bar{\varepsilon})\alpha_1 \frac{R_0 + R - \beta L}{2p} \). For \( \alpha_1 \in [0, \bar{\alpha}] \), \( \sqrt{E_1} \leq 1 - \frac{c}{(\varepsilon - \varepsilon)B} \) and setting \( e = \bar{\varepsilon} \) is a best response for the manager. Now suppose \( e = \varepsilon \). Creditor 1 sets
\[ \sqrt{E_1} = (1 - \varepsilon)\alpha_1 \frac{R_0 + R - \beta L}{2\rho}. \] For \( \alpha_1 \in [\alpha, 1] \), \( \sqrt{E_1} \geq 1 - \frac{c}{(\varepsilon - \epsilon)B} \), and setting \( \varepsilon = \epsilon \) is a best response for the manager.

Increasing the stake of the active monitor increases his monitoring intensity and thereby the probability that he finds the improvement \( R \). The high continuation probability in turn reduces the manager's effort. We find again the result that ex-post efficiency may be at the expense of ex-ante efficiency.

For simplicity, in the intermediate region \([\underline{\alpha}, \overline{\alpha}]\), we will select the equilibrium in which \( \overline{\epsilon} \) is implemented.\(^{10}\)

**Proposition 9** The optimal implementation of \( \overline{\epsilon} \) requires \( \alpha_1^* = \overline{\alpha} \).

**Proof** By Lemma 8, \( \varepsilon = \overline{\epsilon} \) for \( \alpha_1 \in [0, \overline{\alpha}] \). The firm value for \( \varepsilon = \overline{\epsilon} \) is given by

\[
\Pi(\overline{\epsilon}) = \overline{\epsilon} f(L) + \alpha_1 \left\{ (1 - \overline{\epsilon}) \left[ \frac{R_0 + R - \beta L}{2\rho} \right] \right\} + (1 - \overline{\epsilon}) \beta L - \rho [\alpha_1 (1 - \overline{\epsilon}) \frac{R_0 + R - \beta L}{2\rho} ]^2.
\]

As \( \Pi(\overline{\epsilon}) \) is increasing in \( \alpha \) the incentive-compatibility constraint is binding, and \( \alpha_1 = \overline{\alpha} \).

For a given effort level, the firm value is increasing in the amount of monitoring chosen by creditor 1. However, to ensure that the manager chooses the high effort level, the continuation probability must be low enough, i.e. not exceed \( 1 - \frac{c}{(\varepsilon - \epsilon)B} \). Hence it is optimal to set \( \alpha_1 = \overline{\alpha} \) when implementing \( \overline{\epsilon} \). This result is similar to Dewatripont and Tirole (1994) who show that it is optimal to split returns between claimholders, since being residual claimant of the whole profit induces managerial slack. A direct implication of the result above is the following.

**Corollary 4** The optimal implementation of \( \overline{\epsilon} \) may require multiple creditors.

**Proof** By the Proposition above, the optimal implementation of \( \overline{\epsilon} \) requires to set \( \alpha_1 = \overline{\alpha} \). Substituting the lower boundary of \( \rho \) (given in Assumption 10) into the definition of \( \overline{\alpha} \) shows that for the lower admissible value of \( \rho, \overline{\alpha} < 1 \). By continuity,

\(^{10}\)This assumption is innocuous and made only in order to clarify the discussion.
\( \bar{\alpha} \text{ is also smaller than 1 for } \rho \text{-values greater than the lower boundary but in its neighborhood.} \) 

When the active monitor is very efficient, it is necessary to reduce his stake below 1. Otherwise monitoring efforts result in a continuation probability which exceeds the incentive-compatible level and hence destroys the manger's incentives to exert \( \bar{e} \). On the other hand, when the active monitor is not very efficient, there is no reason to introduce additional creditors.

In this section, we have emphasized the gains from allocating the claims among several creditors. The resulting free-rider problem leads to an underprovision of monitoring. This is beneficial to the extent that it prevents renegotiation, i.e. continuation instead of liquidation, and thereby induces the manager to choose \( \bar{e} \). Notice that the optimal implementation of \( e \) requires one creditor only. The reason is intuitive: absent managerial incentives problems monitoring is unambiguously beneficial. Hence, there is no reason to deviate from the arrangement that ensure the highest possible level of monitoring.

How do these results compare with the ones found in the collateral model? Multiple creditors as an instrument to relieve the soft budget constraint problem unambiguously dominate the use of collateral. Collateralization entails a distortion in the input-mix, while multiple creditors avoid this dead-weight loss. However, this conclusion hinges upon the specific features of our model. In particular, when one takes the costs associated with the dispersion of claims into account, the comparison may be reversed.

One can think of several such costs. Having multiple monitors might lead to a duplication of monitoring efforts. Although this does not happen with the monitoring technology chosen in the present model, (unless some creditors have equal shares), in general, several (large) creditors might monitor actively, leading to a (partial) duplication of monitoring costs. These costs can limit the appeal of multiplying the number of creditors.

Moreover, establishing a credit relationship may involve fixed costs (administra-
tive costs, advertising...) which are not proportional to the size of the creditors. For instance, Emmerick and White (1992) report that many companies are unlikely to subject themselves to the registration process and ongoing reporting created by issuing public securities. Additionally, many small public companies would not even be tempted to consider a public offering because of the high fixed costs of a registered offering.

Thus, when the costs associated with the dispersion of claims are introduced, e.g. fixed costs of establishing a borrower-lender relationship, collateral is not necessarily a redundant instrument to tackle the soft budget constraint problem. For example, suppose the optimal \(\alpha\) is very low. Then the firm needs a large number of creditors in order to reduce the stake of the largest creditor. The (fixed) costs required to disperse the claims may then exceed the dead-weight loss due to the use of collateral. Hence, our model does not imply an unambiguous ranking between using the collateral or multiple creditors as means of reducing the monitoring of creditors.

2.6 Different classes of claims: the role of seniority

In the previous section, we assume that the entrepreneur can issue only one class of claims. The concentration of claims is then the only instrument available to fine-tune the incentives to monitor which in turn determined the manager's incentives to exert effort. We now investigate the usefulness of issuing several kinds of debt claims, which have different priority in the event of liquidation. Seniority plays a complementary role to concentration of claims. More precisely, we show that it can be optimal to give seniority to large (monitoring) creditors. In our model, a creditor is senior if he obtains a larger fraction of the in liquidation proceeds than of the continuation returns. Being more protected in bankruptcy, senior creditors are less concerned about the firm being in financial distress. Giving seniority to the monitoring creditors may seem to contradict the intuition we have developed in the previous section. On the
one hand, concentration is useful in providing incentives to monitoring; on the other hand seniority serves the purpose of reducing the amount of monitoring. Given this seeming contradiction, what is the purpose of being concentrated in first place? Why are banks typically senior and secured claimants when their role is to monitor?

Suppose that there are different states of nature after default, and that the optimal implementation of the high effort requires varying degrees of monitoring in each state. If these states were verifiable, every (optimal) level of monitoring could be implemented via a state contingent concentration of claims. Any further instruments would be redundant. However, when the states are not contractible, additional instruments (we will consider seniority) can help to fine-tune the different degrees of monitoring, even if they are substitutes within each state. Indeed, they need not be substitutes across states. On the one hand, in states where liquidation is clearly dominated, seniority plays no role, while it reduces the incentives to monitor in states where liquidation is a valid option. On the other hand, concentration of claims on returns in continuation increases coeteris paribus the extent of monitoring. Giving seniority \(^{11}\) to large claimholders reduces their monitoring in those states where liquidation is desirable, without affecting monitoring in the other states. Summing up, concentration of claims is needed to benefit from monitoring in those states where managerial incentives are not a problem. The priority of monitoring claimholders in bankruptcy is needed to commit to liquidation in those states where continuation would, otherwise, hamper managerial incentives.

To account for the different states of nature and for the multiple classes of debt claims, we extend the previous model as follows.

At date -1, the entrepreneur issues debt contracts in different classes. In particular, claimholder \(i\) obtains a fraction \(\alpha_i\) of returns in continuation and \(\phi_i\) in liquidation. Assume \(\alpha_1 > \alpha_2 > \ldots > \alpha_n\). A creditor is senior if \(\alpha_i < \phi_i\).

In the event of default at date 1, either of two different states of nature can

\(^{11}\)Obviously seniority is only one possible device to relax the soft budget constraint. For instance, in the context of the collateral model of the previous sections, the number of lenders was a possible instrument used in order to reduce the value of continuation, through the free-riding in monitoring.
occur, denoted by $\bar{s}$ and $s$. State $\bar{s}$ occurs with probability $p$ and state $s$ with the complementary probability $1 - p$. The states are not verifiable.

Monitoring takes place at date 0 simultaneously to the manager's effort choice. In contrast to the previous section, monitoring involves two separate activities, each of them is potentially productive only in one of the two states. Creditor $i$ can exert monitoring efforts $E_{i}^{\bar{s}}$ and $E_{i}^{s}$ at a personal cost $\rho(E_{i}^{\bar{s}} + E_{i}^{s})$. The creditors' joint monitoring effort is $E^{w} = \min\{1, \sum E_{i}^{w}\}$, $w = \{s, \bar{s}\}$. In state $w = \{\bar{s}, s\}$, continuation yields a safe return $R_{0}^{w}$ and an additional return $R^{w}$ with probability $\sqrt{E^{w}}$. For simplicity, liquidation yields the same return $\beta L$ in both states.

The assumption that there are no spillovers between $E^{\bar{s}}$ and $E^{s}$ may seem restrictive. However, it should be understood as a simplification of a more general formalization where the two monitoring activities involve the acquisition of information that is only partially overlapping. When a creditor investigates the options that the firm has in one state, he will also learn something about the firm's option in the other state. Nevertheless, these are pieces of information which will be of use only in one state of nature. One argument relies upon the feature that there are only partial spillovers between the two monitoring activities. It does not require that $E^{\bar{s}}$ and $E^{s}$ are completely separate activities. This is merely a convenient simplification.

Finally we need to amend Assumptions 10 and 11 to the extended framework with two states of nature. To keep the exposition simple, we assume that monitoring is too costly to guarantee that the improvement in the continuation value $R^{w}$ is found with probability 1 in either state.

**Assumption 12** $\rho \geq \frac{(1-\epsilon)pR^{\bar{s}}}{2} > \frac{(1-\epsilon)(1-p)[R_{0}^{s} + R^{s} - \beta L]}{2}$

To avoid that continuation is always the better option despite the high monitoring costs we impose additional restrictions.

**Assumption 13** $R_{0}^{\bar{s}} > \beta L > R_{0}^{s}$, $R_{0}^{s} + R^{s} > \beta L$

Unless inefficient renegotiation occurs, the firm is always continued is state $\bar{s}$, while continuation in state $s$ requires that monitoring has been successful. For the
moment we assume efficient renegotiation. Assumptions 12 and 13 capture the idea that liquidation is a viable option in some states of the world, while in other states there is a clear incentive to keep the firm as a going concern. In the latter case, monitoring is unambiguously beneficial, while in the former case it might destroy the manager’s incentives.

We start the analysis by characterizing the equilibrium amount of monitoring.

**Lemma 10** Creditor \( i \) provides monitoring relevant for state \( \bar{s} \) iff \( \alpha_i = \max\{\alpha_k\} \).

Creditor \( j \) provides monitoring relevant for state \( \underline{s} \) iff \( \alpha_j(R_0^\underline{s} + R^\underline{s}) - \phi_j\beta L = \max\{\alpha_k(R_0^\underline{s} + R^\underline{s}) - \phi_k\beta L\} \)

**Proof** Creditor \( k \)'s program in direction \( \bar{s} \) is

\[
\max p(1 - e)[\alpha_k(R_0^\bar{s} + R^\bar{s})\sqrt{E^\bar{s}} + \alpha_k R_0^\bar{s}(1 - \sqrt{E^\bar{s}})] - \rho E_k^\bar{s}
\]

Similarly, in direction \( \underline{s} \), the program is

\[
\max (1 - p)(1 - e)[\alpha_k(R_0^\underline{s} + R^\underline{s})\sqrt{E^\underline{s}} + \phi_k\beta L\bar{s}(1 - \sqrt{E^\underline{s}})] - \rho E_k^\underline{s}
\]

The rest of the proof follows from the proof of Lemma 8. \( \square \)

Given that \( \alpha_1 > \alpha_2 > \ldots > \alpha_n \) by assumption, creditor 1 is the only active monitor in direction \( \bar{s} \). That is, only creditor 1 gathers information that is relevant for the continuation payoff is state \( \bar{s} \). As regards monitoring in direction \( \underline{s} \), creditor 1 is not necessarily the only active monitor, nor is there necessarily only one active monitor. The condition for being an active monitor in state \( \underline{s} \) might hold for other creditors than creditor 1 and might hold for several creditor simultaneously. Suppose that creditor 1's claim on the continuation return \( \alpha_1 \) is small relative to his claim on the liquidation proceeds \( \phi_1 \). If there is a junior creditor whose stake in the continuation payoff is sufficiently large, he will be the active monitor in direction \( \underline{s} \).

To fix ideas we assume that creditor 1 monitors in both directions and all other creditors free-ride. This assumption implies that there is a single large creditor,
henceforth the bank, and many small ones, henceforth the bondholders.

**Assumption 14** \( (\alpha_1 - \alpha_j) \frac{(R_0^k + R_0^u)}{\beta L} > (\phi_1 - \phi_j) \) for all \( j = 2, \ldots, n \).

Assumption 14 implies that the two conditions for being an active monitor hold only for creditor 1. In particular, it excludes the presence of a second large creditor who is junior. In view of the stylized facts, this assumption is not very restrictive. In fact, when a firm is funded by several large creditors (banks), they are often part of a syndicate with equivalent claims. For our purposes, they can be viewed as one large creditor.

**Lemma 11** Given \( e \),

- \( E_1^T \) is increasing in \( \alpha_1 \) and independent of \( \phi_1 \);
- \( E_1^L \) is increasing in \( \alpha_1 \) and decreasing in \( \phi_1 \).

**Proof** By Lemma 10 the efforts are

\[
\sqrt{E_1^T} = \frac{\alpha_1 (1-e)pR^s}{2\rho} = \frac{(1-p)(1-e)[\alpha_1(R_0^k + R^u) - \phi_1 \beta L]}{2\rho}
\]

While both types of monitoring are increasing in the stake of the active monitoring, \( E_1^T \) is independent of the liquidation return, while \( E_1^L \) is decreasing in \( \phi_1 \). A high fraction of the liquidation value reduces monitoring effort in one direction without affecting monitoring in the other direction.

**Lemma 12** If \( p > 1 - \frac{c}{(e-\xi)e} \) the high effort cannot be implemented. The optimal debt design is \( \alpha_1^* = 1 \) and \( \phi_1^* = 1 \).

**Proof** Given that renegotiation is efficient and \( R_0^c \geq \beta L \), the project is always continued in state 3. If \( p > 1 - \frac{c}{(e-\xi)e} \) liquidation after default occurs with a
probability less than $\frac{c}{(\varepsilon - \underline{\varepsilon})B}$. By Lemma 3 the low effort is chosen. By Lemma 11 the firm value as a function of $\alpha_1$ and $\phi_1$ is equal to

$$\Pi(\alpha_1, \phi_1) = \varepsilon g f(L) + (2\alpha_1 - \alpha^2_1)\frac{|p(1-p)|R_{\bar{s}}^2}{4\rho} + (1-\varepsilon)[pR_{\bar{s}}^* + (1-p)\beta L] + [\frac{(1-p)(1-\varepsilon)}{4\rho}]^\prime \alpha_1 (R_{\bar{s}}^* + R_{\bar{s}} - \phi_1 \beta L)]((2 - \alpha_1)(R_{\bar{s}}^* + R_{\bar{s}}) - (2 - \phi_1)\beta L).$$

$\Pi(\alpha_1, \phi_1)$ is decreasing in $\phi_1$, and for $\alpha_1 \leq \phi_1$ increasing in $\alpha_1$. As $\frac{\partial \Pi(\alpha_1, \phi_1)}{\partial \alpha_1} > \frac{|\partial \Pi(\alpha_1, \phi_1)|}{\partial \phi_1}$ for all $\alpha_1, \phi_1 \in [0, 1]$, given $\alpha_1 \leq \phi_1$, it is optimal to set $\alpha_1 = \phi_1 = 1$. □

Since $R_{\bar{s}}^* > \beta L$, continuation dominates liquidation in state $\bar{s}$, irrespective of the success creditor 1 has in finding the continuation improvement $R_{\bar{s}}^*$. When the probability of state $\bar{s}$ exceeds the incentive compatible level of continuation $1 - \frac{c}{(\varepsilon - \underline{\varepsilon})B}$, the manager always shirks. Absent managerial incentives considerations, a fully concentrated ownership structure is optimal because it produces the maximum amount of monitoring, avoiding the free-rider problem in monitoring. In what follows, we assume that the high effort can be implemented.

**Assumption 15** $p < 1 - \frac{c}{(\varepsilon - \underline{\varepsilon})B}$

**Lemma 13** If $p + (1 - p)(1 - \varepsilon)\frac{R_{\bar{s}}^* + R_{\bar{s}} - \beta L}{2\rho} < 1 - \frac{c}{(\varepsilon - \underline{\varepsilon})B}$ the high effort can be implemented and the optimal debt design is $\alpha_1^* = 1$ and $\phi_1^* = 1$.

**Proof** With $\alpha_1 = 1$ the continuation probability is at its maximum. For $p + (1 - p)(1 - \varepsilon)\frac{R_{\bar{s}}^* + R_{\bar{s}} - \beta L}{2\rho} < 1 - \frac{c}{(\varepsilon - \underline{\varepsilon})B}$ the resulting complementary liquidation probability induces the manager to choose $\varepsilon$ for all $\alpha_1 \in [0, 1]$. As $\varepsilon$ enters $\Pi(\alpha_1, \phi_1)$ only as a scaling factor, the optimality of $\alpha_1 = \phi_1 = 1$ follows from Lemma 12. □

When monitoring is sufficiently costly and the probability of state $\bar{s}$ is sufficiently low, managerial incentives are not endangered even by the maximum possible level of monitoring. We study now the more interesting case where the manager’s incentives are affected by the extent of monitoring undertaken in state $s$.

**Assumption 16**

$$p + (1 - p)(1 - \varepsilon)\frac{R_{\bar{s}}^* + R_{\bar{s}} - \beta L}{2\rho} > 1 - \frac{c}{(\varepsilon - \underline{\varepsilon})B}$$
Under Assumption 16 implementation of $\bar{\varepsilon}$ requires to depart from $\alpha_1 = \phi_1 = 1$ since this security structure induces a continuation probability which is not incentive-compatible.

**Proposition 10** Under Assumption 16 the optimal implementation of the high effort requires $\alpha^*_1 < 1$ and $\phi^*_1 = 1$.

**Proof** The constraint on $E^z$ is binding so that

$$p + (1 - p)(1 - \bar{\varepsilon}) \frac{\alpha_1 (R^u_0 + R^u) - \phi_1 \beta L}{2 \rho} = 1 - \frac{c}{(\bar{\varepsilon} - \varepsilon) B}$$

But this leads to undersupply of $E^\bar{z}$. If $\phi_1 < 1$, a joint variation of $\alpha_1$ and $\phi_1$ such that

$$\Delta \phi = \frac{R^u_0 + R^u}{\beta L} \Delta \alpha$$

($\Delta \alpha > 0$) is neutral for the constraint on $E^z$, but by Lemma 13 increases firm value. □

**Corollary 5** The bank is senior vis-à-vis bondholders.

This is the main result of this section. It is optimal to allow active monitors to be senior in liquidation with respect to non-monitoring creditors. Seniority together with concentration of claims helps to fine-tune the incentives in different states of the world. It is possible to obtain the benefits of monitoring, i.e. ex-post efficiency without paying (or paying at a lower extent) the costs of monitoring, i.e. ex-ante inefficiency.

We have assumed so far that renegotiation is efficient, and hence that the project is always continued in state $\bar{z}$. However, several authors argue that the dispersion of claims leads to inefficiencies in renegotiation. If $\alpha_1 R^u_0 < \phi_1 \beta L$, the inability of bondholders to renegotiate their claims so as to compensate the bank will lead the latter to liquidate inefficiently. Since the solution we have derived in Proposition 10 does not exclude the possibility that $\alpha_1 R^u_0 < \phi_1 \beta L$ inefficient renegotiation affects the optimal debt design.
There are two ways to deal with this problem. The first is to restore efficient, or less inefficient, renegotiation. This can be done by decreasing the stakes of the bank. Indeed it is observed in practice that firms tend to borrow from a main bank and smaller stakes from other banks. This solution entails the cost of giving up part of the benefits deriving from having concentrated claims. Second, reduce the protection of the bank in liquidation, i.e. $\phi_1 < 1$. This is also costly since $\alpha_1$ has to decrease in order to keep management under a tight incentive scheme. The cost of this second solution is that $E^g$ will be reduced. But this cost may be lower than the cost of having inefficient liquidation in the good state. Moreover, it is commonly observed that banks are not entitled to all the proceedings of liquidation. While this fact is generally attributed to legal constraints (the payoff in liquidation cannot exceed the face value of claims), this feature can be explained by our theory.

Both means to alleviate the threat of inefficient renegotiation lead to a reduction in the amount of monitoring exerted in equilibrium.

Our model of seniority and the model of collateral developed in the previous sections have a common theme. The underlying idea is to increase the liquidation payoff for monitoring parties. Liquidation becomes relatively more attractive for the monitors, thereby reducing their monitoring efforts. This results in higher incentives for the management ex-ante. Both models are based on the idea that liquidation should, in some states, be an attractive option for the creditors whose monitoring effort could lead to excessive softness after financial distress.

In our framework, information acquired ex-ante may trigger efficient renegotiation ex-post. The contract theory literature has emphasized the ex-ante inefficiency associated with ex-post renegotiation. As a consequence, we obtain the implication that it may be better for the claimholders to commit not to acquire too much information. This is achieved via the design of securities and their ownership structure. In particular, it is possible to exploit a free-rider behavior induced by the dispersion of claims: ex-post efficiency being a public good among claimholders, small claimholders will free-ride on its provision, i.e. on the acquisition of information ex-ante.
Notice that it is the concentration of claims ex-post which provides the incentives to acquire information ex-ante. Hence, the robusteness of our mechanism relies on the ability to commit to a (relatively) dispersed ownership structure ex-post. This is not problematic in our framework. Indeed, once monitoring has been realized, there is no scope for profitable trade among claimholders. Therefore, the ex-ante concentration is maintained ex-post. In particular, a dispersed ownership structure of claims ex-ante constitutes in itself a commitment to a dispersed ownership ex-post.

As a consequence, our scheme does not rely on the inability to concentrate claims ex-post. Such an inability could rely on a different free-rider problem, identified by Grossman and Hart (1980). Suppose that claims are dispersed and that concentration would improve their value. Each atomistic individual claimholder will not tender his claims at a price that does not fully incorporate the improvement. Hence, an agent bidding for claims so as to concentrate them will get little of the surplus. As a result, concentration will be undersupplied. Also, our scheme does not rely on similar inefficiencies in ex-post bargaining due to the large number of agents or to information asymmetries among bargainers. Indeed, we assume that ex-post bargaining is efficient given the information available to the agents.

2.7 Concluding Remarks

In this paper, we try to give a possible explanation for the counterintuitive observation that financial institutions that are supposed to monitor more efficiently are actually more protected in bankruptcy than the dispersed claimholders that are not supposed to perform any monitoring. A high liquidation value for the monitors is a means of obtaining the benefits of monitoring without bearing the costs of monitoring.

We focus only on collateral and seniority as instruments to reduce the soft bud-

---

12 This is irrespective of whether the large creditor's information about the intensity or the outcome of monitoring is private or not.
13 Grossman and Hart (1980) developed their argument in the context of tender offers.
14 This would be the case with ex-post monitoring.
15 In that, our approach is different from Zingales (1994) and other bargaining based theories.
get constraint problem. Obviously other arrangements as covenants could work in the same direction. In particular, we have omitted considering other mechanism of obtaining concessions from the management. Form this point of view, our theory is quite close in spirit to the work of Dewatripont and Tirole (1994). In their model control rights are the variable which determines the intensity of monitoring. In our model, as well as in Crémer (1993), we assume as given control rights (each party can liquidate in case of distress unless renegotiation remove this threat) and we focus our analysis on different splits of monetary returns. An interesting extension of our model would be the joint endogenization of control rights and return stream. The analysis of these and other exciting questions must await future research.
Chapter 3

Decentralization, Mobility Costs, and the Soft Budget Constraint
(joint with F. Daveri)

3.1 Introduction

The soft-budget constraint syndrome [...] pertains whenever a funding source - e.g. a bank or government - finds it impossible to keep an enterprise to a fixed budget. That is, the enterprise can extract ex-post a bigger subsidy or loan than would have been considered efficient ex-ante. As Kornai (1979) forcefully demonstrated, the centralized economies of eastern Europe were rife with soft budget constraint. By contrast, the more decentralized economies of the West have seemed far less prone to the syndrome, although they are by no means immune. (Maskin, 1994, p.1)

The soft budget constraint syndrome is indeed a widespread one on a worldwide scale and a society's resource allocation system has a bearing on its extent and diffusion. The bulk of the recent literature on the subject has emphasized the effectiveness
of decentralized decision-making in implementing a hard budget constraint.\footnote{One notable exception regards decentralization of monetary policy in Qian and Roland (1994).}

While decentralization has been attached a variety of different meanings, the common underlying idea is that, through decentralization, the burden of some decision is split between multiple parties, each of whom does not fully internalize its consequences.\footnote{See Dewatripont and Tirole (1995) on this topic.}

In Dewatripont and Maskin (1995), decentralization amounts to letting a tougher party take the decision about refinancing the project of a firm. Their model studies the credit relations between a firm and several banks sequentially interacting with the firm, each at a different point in time. By assumption, the bank extending the credit at the outset is endowed with too limited an amount of resources to be able to provide the firm with refinancing (in case the firm’s project is slow). In turn, absent information sharing, the bank coming next has coarser information than the former one in the sequence. As a result, the extractible surplus from continuation of the project shrinks when two banks are involved rather than one. This restraints the firm’s ability of borrowing, thus hardening its budget constraint.

Another reason why having multiple parties involved in the decision-making process is beneficial is suggested by Dewatripont and Tirole (1994). Control and return rights naturally divide holders of claims in two distinct classes. Bondholders tend to favor less risky courses of actions (like liquidation) than equity holders, who would rather favor continuation of the firm project, if the risky return prospects are bright enough. Then control may be optimally allocated to bondholders in unfavourable states of nature, where they can extract more concessions from the managers of the firm. Note that the difference in attitudes towards refinancing the enterprise stems from the nature of the claims rather than from asymmetric information as in Dewatripont and Maskin’s paper.

A third approach to the soft budget constraint problem, adopted by Burkart et al. (1994), asks whether the tendency for small claimholders to free ride on monitoring efforts can be exploited to achieve an appropriate reshaping of firm incentives. When
claims are dispersed, financial distress will be more often associated to liquidation and refinancing will become more unlikely.\footnote{The papers mentioned above are by no means an exhaustive list of the literature on decentralization and soft budget constraint. See Maskin (1994) for a more extensive treatment.}

All in all, having a single party which internalizes all of the consequences of a certain decision can induce wrong ex-ante incentives. The proposed remedy is delegation of the decision to multiple parties, each of whom internalizes only part of the consequences of the action in its objective function.

Obviously an important determinant of the impact that decentralization has on the hardness of the budget constraint is the stake of each of the parties involved. In our model the crucial variable is the presence of workers’ mobility costs. More precisely, the problem we analyze is the one faced by a regional government in charge of either refinancing or closing a loss-making firm, whose continuation has, however, a positive social value for (a fraction of) the citizens of the region itself.

Regional welfare includes profits as well as citizens’ utility. Individual utilities are affected positively by labor earnings and the quantity consumed of a congestible public good. Shutting down the firm implies a social loss\footnote{This can be explained in many ways. As Segal (1994) points out, one leading case is when the firm has monopsony power in the labor market. Then workers can have rents and quasi-rents that would be dissipated in case the firm is shut down. The origin of workers’ rents can be due to inability of perfectly discriminate workers on the basis of their characteristics or to the presence of unions or simply in order to induce high effort from them.} as long as the rents earned by the firm employees cannot be recouped anywhere within the same region. In common parlance, the firm (or sector) is ‘strategic’ for the region. Chrysler was strategic for Detroit in the early eighties and the Sulcis mines were strategic for Sardinia in 1994.

In this framework, regional out-migration of the prospective unemployed helps ‘export’ (or share) the burden of the bail-out decision on some other region’s shoulders. Once migration is allowed, a situation where multiple parties are involved in the decision over the budget of the enterprise is achieved, even though at the beginning only one party was facing the choice problem. In this framework, mobility costs determine the stake of each party in the decision process and to what extent decentralization is beneficial in solving the soft budget constraint problem.
We believe that the attribution to mobility costs of a role in the stake allocation is a genuinely new feature of our model. In Dewatripont and Maskin (1995), exogenous constraints on credit capacity cause the refinancing decision to be delegated to a tougher party. In Dewatripont and Tirole (1994), the initial contract determines the states (and the stakes) of different claimholders. In our paper, as a result of labor mobility, the stake of each party (region) in the refinancing process is endogenously determined and depends upon such things as the costs of mobility and the number of job openings in the region of immigration.

The distinguishing feature of decentralization with respect to centralization is the fact that one region (region B in our model) faces higher marginal costs in transferring money to the firm than the region where the firm is localized (region A in our model) and the State. We will model this higher cost by a moral hazard problem existing between the two regions. An equivalent model of hidden information could deliver the same conclusions. In the same vein as Dewatripont and Maskin (1995), we will suppose that region B has a coarser information about the firm than region A. An informational rent accrues in this case to region A.  

Region B has to pay the transfer itself and the informational rent. The State, instead, fully internalizes the loss borne by region B without encountering asymmetric information shortcomings. In the case of positive externalities created by refinancing the firm, the decreased availability of funds for refinancing has a beneficial effect on the tightness of ex-ante incentives.

---

5The underlying idea of this model is that information about the firm is always available at no cost to the local authority. Moreover, in the case of centralization, we assume that there are no agency problems inside the State, so that even if an agency of the State (say, a Préfecture in France) collects local information, it has no incentives not to report honestly to the central authority.

In our model region A and the State differ only in their objective function and not in the information they possess. Obviously this in an unrealistic assumption. However, the literature, Dewatripont and Tirole (1995) being a notable exception, has not provided yet a theory of the incentives to collect information. Do local authorities have superior information with respect to central authorities? If so, is that because of better technology or superior incentives? Who collects the information and who should instead: politicians, bureaucrats? Are the incentives of bureaucrats aligned with those of politicians?

These problems also arise in the literature on the boundaries of the firm. The moral hazard problem that we borrow from Dewatripont and Maskin (1995) would hold also between different departments of a bank rather than between two different banks. In this sense a theory of decentralization should be based on a more solid theory of the boundaries of the firm.
On the other hand, in case of negative externalities, centralization implies a smaller transfer to the firm and, thus, a harder budget constraint than in the decentralized case.

These results bear some similarity with those obtained by Klibanoff and Morduch (1993) about the relevance of the size of externalities. If regions enjoy some informational advantage over the government, this adds to the costs of coordination/centralization. If the gains from coordination are small (i.e. the size of the externality is small), then decentralization may be preferable. A decentralized system delivers better welfare results, if the externality-induced loss is more than offset by savings on informational rents. Symmetrically, centralization is more likely to be preferable when externalities are large. There is however an important difference with their model: in their model asymmetric information has no positive social effect. A world of complete information would always be preferable. In our model this is not the case: informational asymmetries have the beneficial effect of reducing the softness of the budget constraint.

Our paper can also be compared with Qian and Roland (1994). They find that while fiscal decentralization can be useful in implementing a hard budget constraint, monetary decentralization actually softens the budget constraint. We offer a reinterpretation of their results in terms of positive and negative externalities among regions.

But, as already pointed out, our model has also a feature which is not present in the models above: the nature of the externality (i.e. whether it is positive or negative) depends on the mobility cost of workers. We can then study how the process of integration inside a federation and the process of decentralization interact. Our main finding is that decentralization is efficient in implementing a hard budget constraint only if mobility costs are low enough. In other words, decentralization without labor mobility may fail to deliver tougher budget discipline. Hence we reach the conclusion that the degree of 'similarity' among different regions in a federation is an important factor in determining the welfare outcomes of fiscal federalism.
The paper is organized as follows. In section 2, the general set-up for a discussion of the soft budget constraint problem when externalities across regions exist is formalized. Section 3 develops the migration example, deriving the equilibrium relation between migration and its determinants. Sections 4 and 5 show how mobility costs affect the relative desirability of decentralization and centralization in decision-making. Section 6 presents an extension which we believe of some interest, the case of multiple host regions. Section 7 offers some elements to frame our discussion within the debate on the positive theory of the size of the government. Section 8 concludes.

3.2 Externalities and the soft budget constraint

In this section we compare how two alternative regimes of decision-making (centralization and decentralization) fare in softening or hardening the firm budget constraint in the presence of externalities among regions. Our model builds upon Dewatripont and Maskin (1995). A firm must decide on the amount of a cost-reducing investment. High levels of investment guarantee that the firm will be profitable, while low levels of investment potentially lead the firm to a situation of financial distress. In the latter case, however, the firm may be bailed out by a statal or regional authority able to provide the firm with the necessary resources to continue activity. This outcome is socially inefficient, but privately preferable for the firm’s management. The crucial determinant of whether investment will be high or low is the probability that the firm will be refinanced ex-post. The authority in charge of refinancing the firm would like to be able to commit not to refinance. However, in these circumstances, this is known to be dynamically inconsistent, for, once the investment decision is taken, it becomes socially optimal to bail-out the firm. The reasons for statal or regional bail-outs are only too many. We focus on one which seems to us a pervasive concern, i.e. the public distaste for the consequences of unemployment generated by the firm’s closure.

In a centralized setting, the two-region State\footnote{It is worth pointing out that in our version of centralization, regions have no active role.} we are dealing with is in charge for the decision of replenishing the losses of the firm in financial distress. The State is
benevolent and fair, so that its utility is simply the sum of the utilities of the two regions and finances the transfer to the firm by raising the level of corporate taxation\textsuperscript{7}. The State has complete information about the firm.

In a decentralized setting, both regions non-cooperatively determine how much of the regional budget is to be allocated to bailing-out the firm. The distinguishing feature of decentralization is that region B faces a moral hazard problem with respect to region A and hence has a higher marginal cost of making transfers to the firm. The idea we have in mind is that region B has access to coarser information about the firm with respect to the state or to region A, in the same spirit as Dewatripont and Maskin (1995).

The main conclusion we derive in this section is that the sign of the externality is the crucial variable to assess which regime implements a harder budget constraint. In particular, we prove that decentralization is more effective in implementing a hard budget constraint only if positive externalities among regions exist. In the opposite case, centralization is preferable.

3.2.1 The model

The overall set-up of the model is borrowed from Dewatripont and Maskin (1995). A managerial firm has a project. The returns to the project depend on managerial investment effort, which can take two values: \( e \in \{e, \bar{e}\} \). The additional disutility of effort \( \bar{e} \) for the manager is \( K > 0 \). If \( \bar{e} \) is chosen, the firm is viable for sure and pays a return of \( R \) to its shareholders and a private, non-taxable benefit of \( B_q > 0 \) to the managers. If \( e \) is chosen, the firm will be in distress and it will need external funding in order to remain active. Investing resources \( \rho \frac{B_q^2}{2} \) into the firm, the firm becomes again viable with probability \( E \). In this case, the firm pays no return to its

\textsuperscript{7}In order to bail-out the firm, we assume that only profits of other firms are taxed. Since capital is (imperfectly) mobile this creates the distortion. We further assume that if capital migrates from one region, it goes abroad and not in the other region. The case of regional competition for capital is studied in detail in Qian and Roland (1994).
shareholders, but a private benefit of \( B_q > 0 \) accrues to the manager. We assume

\[
B_q - K < B_s.
\]

If the firm is refinanced with certainty in case of financial distress, the manager will shirk. With probability \( 1 - E \) the firm must be closed and neither monetary return nor private benefit arise. We rule out for simplicity, as in Dewatripont and Maskin (1995) and Qian and Roland (1994), monetary incentives for the manager. The underlying idea is that private benefits are very high so that aligning incentive through the use of monetary bonuses is too costly for the firm's shareholders. The crucial ingredient of the model is obviously the presence of private benefits and not the absence of monetary incentive schemes for the manager.

In case the firm is closed, a social loss \( l \geq 0 \) is caused to region A, where the firm is located, and an externality \( b \) (positive or negative) to region B. While we take \( l \) and \( b \) as given in this section, we will argue that a plausible interpretation for \( l \) is, as in Segal (1994), the loss of workers' rents. The externality \( b \) is amenable to different interpretations: in the next section, we will pick one and show how migration of labor, potentially unemployed in region A when the firm shuts down but possibly employed in region B, may produce an externality of either sign on region B.

In our framework, shareholders do not have incentives to refinance the firm (profits are zero in any case), so that only the state or the regions can bail-out the firm.

Consider first the manager's problem. The manager takes \( E \), the probability that the firm is refinanced ex-post, as given and chooses \( \tilde{e} \), the socially efficient effort, only if he finds it privately preferable, i.e. only if

\[
B_q - K \geq EB_s,
\]

or

\[
E \leq \frac{B_q - K}{B_s} = \tilde{E}.
\]

If \( E \leq \tilde{E} \), a hard budget constraint outcome will emerge; if the inequality is reversed,
a soft budget constraint outcome will be implemented.

Before comparing the relative merits of centralization and decentralization, we need to establish that bailing-out the firm is always socially optimal ex-post, even taking into account the externality on region B. Then we assume:

**Assumption 17** \( l + b \geq 0 \)

that sets a lower bound for \( b \).

Under financial distress, the State faces the following problem

\[
\max_{E} -(l + b)(1 - E) - \rho(1 + \lambda) \frac{E^2}{2}
\]

where \( \lambda \) is the shadow cost of public funds, assumed to be non-negative. Raising one unit of public funds costs \( (1 + \lambda) \).

The solution of the problem is

\[
E^C = \frac{l + b}{(1 + \lambda)\rho}.
\]

To ensure an internal solution for \( E \) we assume

**Assumption 18** \( b < \rho(1 + \lambda) - l \)

which sets an upper bound for \( b \). Since \( b \) can take both positive and negative values, the right-hand side of the inequality is certainly positive.

We have then proved the following

**Proposition 11** Centralization implements a hard budget constraint outcome iff

\[
\frac{l + b}{(1 + \lambda)\rho} \leq \bar{E}.
\]

The decentralization case is less straightforward. We assume that region A only can directly invest resources into the firm. Region B can however make a transfer to region A. The transfer cannot be made contingent on the amount of effort \( E \) exercised
by region A, but, as in Dewatripont and Maskin (1995), only on its final outcome, i.e. whether the firm is closed or not. We have then a typical situation where a moral hazard problem between regions exists. The unverifiability of region A’s effort creates an informational rent in favor of region A and makes the cost of transferring funds from region B to the firm higher.

We assume that region B has all the bargaining power and offers a contract to region A. A contract is a pair \( \{s, t\} \): \( s \) is the transfer if the firm remains open and \( t \) is the transfer if the firm is closed. As a convention, we assume that transfers are from region B to region A. We impose limited liability for region A, i.e. \( s, t \geq 0 \).

While this assumption is partly ad hoc and mainly imposed in order to have an interesting moral hazard problem, it can be defended on the ground that for politicians of region A can be hard to justify to the voters a transfer to region B when the firm is closed. Decentralization is then associated in our framework to a regime where only unilateral promises of payment can be enforced. The analysis of decentralization differs depending on whether:

a) \( b > 0 \)

or

b) \( b < 0 \).

In case a), region B has an interest in having the firm open and so it is ready to pay a net positive transfer in case the firm remains active. It will obviously hold \( s \geq t \).

The problem of region B\(^8\) is then

\[
\max_{t, s} \left[ -E(1 + \lambda)(s) - (1 - E)(b + (1 + \lambda)t) \right]
\]

\(^8\)Probably an adverse selection model would better deliver the idea that region A has access to information not available to region B about the firm.

\(^9\)We will assume that the shadow cost of public funds in centralization and decentralization is the same. Though not realistic, this assumption keeps things equal between centralization and decentralization. The State and the regions might face different costs of fund raising. The State is usually endowed with a wider set of instruments than each region alone, but a lower size of the government budget under decentralization may feed into lower marginal cost of funds. This is a very important question to analyze, but we omit a serious discussion of the issue.
subject to the constraints

\[ E = \arg\max \ U^A(t, s) = E(1 + \lambda)s - [l - (1 + \lambda)t](1 - E) - \rho(1 + \lambda) \frac{E^2}{2} \]

and

\[ s, t \geq 0. \]

The subsidy from region B partially relieves region A from the burden of taxation. This explains why the amount of the transfer from B to A is augmented by a factor \( 1 + \lambda \).

Obviously it will hold \( t = 0 \) so that the solution to A’s problem is easily derived:

\[ E^D = \frac{l + (1 + \lambda)s}{\rho(1 + \lambda)}. \]

Substituting for \( E \) the expression found above we have

\[ s = \frac{b - l}{2(1 + \lambda)} \quad \text{if} \quad b \geq l \]

and

\[ s = 0 \quad \text{otherwise}. \]

Substituting \( s \) in the expression for \( E \) we finally have

\[ E^D = \frac{l + b}{2\rho(1 + \lambda)} \quad \text{for} \quad b \geq l \]

and

\[ E^D = \frac{l}{\rho(1 + \lambda)} \quad \text{for} \quad b < l. \]

Summing up, we have

**Proposition 12** If \( 0 < b < l \), decentralization implements a hard budget constraint outcome iff

\[ \frac{l}{\rho(1 + \lambda)} \leq \tilde{E}. \]
If $b > l$, decentralization implements a hard budget constraint outcome iff

$$\frac{l + b}{2\rho(1 + \lambda)} \leq \bar{E}.$$

The main implication of Proposition 12 is the following:

**Corollary 6** If $b > 0$, then $E^C > E^D$, that is decentralization implements a harder budget constraint than centralization.

Suppose for the moment that $E$ is verifiable i.e. that a contract $\{E, v(E)\}$, where $v(E)$ is the transfer associated to $E$ is feasible (or equivalently that there is no limited liability constraint). Then the contract that region B proposes is

$$v(E) = -\frac{b^2 + 2bl}{2\rho(1 + \lambda)} + \frac{b}{1 + \lambda}E$$

and

$$E^D = \frac{l + b}{\rho(1 + \lambda)} = E^C$$

is implemented. Notice that the slope of the incentive scheme is positive: high effort is rewarded.

When instead $E$ is not verifiable, we have an additional effect due to the informational rent that must be paid because of the moral hazard problem. This additional effect goes in the direction of reducing $E^D$: the higher cost of making transfers faced by region B positively enhances fiscal discipline, for the strategic firm is subtracted a part of the rent extractible under financial distress.

In case b), that is when $b < 0$, region B prefers having the firm closed. This makes it willing to pay a transfer $t$ for the firm in region A not to be refinanced. Obviously, in this case, $t \geq s$.

The problem of region B is then

$$\max_{s, t} - E(1 + \lambda)s - (1 - E)[b + (1 + \lambda)t]$$

subject to the constraints
\[ E = \text{argmax } E(1 + \lambda)s - (1 - E)[l - (1 + \lambda)t] - (1 + \lambda)\rho \frac{E^2}{2} \]

and

\[ s, t \geq 0. \]

Since it will hold \( s = 0 \), the solution of A's problem is

\[ E^D = \frac{l - (1 + \lambda)t}{\rho(1 + \lambda)}. \]

The solution of B's problem is

\[ t = \frac{l - b}{2(1 + \lambda)} - \frac{\rho}{2} \text{ if } b < l - \rho(1 + \lambda) \]

and

\[ t = 0 \text{ otherwise}. \]

Substituting back into the expression for \( E \), we find

\[ E^D = \frac{l + b}{2\rho(1 + \lambda)} + \frac{1}{2} \text{ if } b < l - \rho(1 + \lambda) \]

and

\[ E^D = \frac{l}{\rho(1 + \lambda)} \text{ if } b > l - \rho(1 + \lambda) \]

We have then proved

**Proposition 13** If \( b < l - \rho(1 + \lambda) \), a decentralized system implements a hard budget constraint iff

\[ \frac{l + b}{2\rho(1 + \lambda)} + \frac{1}{2} \leq \bar{E}. \]

If \( l - \rho(1 + \lambda) \leq b \leq 0 \), a decentralized system implements a hard budget constraint iff

\[ \frac{l}{\rho(1 + \lambda)} \leq \bar{E}. \]

The main implication of Proposition 13 is
Corollary 7 If \( b < 0 \), \( E^C < E^D \) that is decentralization implements a softer budget constraint than centralization.

In the absence of a moral hazard problem, i.e. when \( E \) is contractible, (or without the limited liability assumption) region B can offer, as before, a payment schedule

\[
v(E) = -\frac{b^2 + 2bl}{2\rho(1 + \lambda)} + \frac{b}{1 + \lambda} E
\]

and

\[
E^D = \frac{b + l}{\rho(1 + \lambda)} = E^C
\]

is implemented. Notice that now the slope of the incentive scheme is negative: high effort is penalized.

When \( E \) is not verifiable, instead, the higher transfer cost due to the moral hazard problem elicits a softer budget constraint, because the value of shutting down the firm is now lower for the region than for the State.

Summing up, our main finding of this section is contained in

**Proposition 14** A necessary (but not sufficient) condition for decentralization to be preferable to centralization as a means to overcome the soft budget constraint problem is that externalities among regions are positive.

Our findings can be compared with Qian and Roland's (1994), where the instrument by which firms in distress are refinanced determines whether positive or negative externalities show up across regions. In case of fiscal decentralization, if a region has a soft budget constraint, it has fewer resources to attract foreign capital. Capital will then flow to other regions, creating a positive externality for them. The externality is however not internalized at the region level. Each region is then tougher than the State, which, instead, also takes into account the beneficial spillover effects on other regions.

Monetary decentralization, instead, creates a negative externality for the benefits from decentralized inflationary financing are fully internalized at the local level, while
its costs are equally shared by all regions. Excessive money creation is the most likely outcome in a decentralized regime of monetary policy.

3.3 Public good provision and migration as sources of externalities

In the remaining part of the paper we endogenize l and b in a model with labor mobility and finally draw some conclusions as to how mobility costs affect the choice between centralization and decentralization as alternative arrangements for decision-making. Labor mobility is not the only conceivable mechanism to generate spillovers across regions. However allowing for labor mobility seemed to us well tuned with our overall modelling strategy, which focuses on one instance of soft budget constraint problem linked to employment concerns.

Suppose that region A incorporates both the profits of firms located within its boundaries and the utilities of its citizens\(^\text{10}\) in its welfare function. Then shutting down the firm implies for the region the loss of the rents accruing to the \(L_F\) workers employed by the strategic firm.

The region would then be ready to pay a transfer to the firm in order to maintain the level of employment unchanged\(^\text{11}\).

However, if workers can freely migrate between regions, the amount of the transfer will crucially depend on the number of workers who decide to migrate to region B rather than staying unemployed in region A. Migration will thus reduce the social loss in region A by exporting part of it into region B. In our framework there are two reasons why the loss is reduced (and therefore the soft budget constraint problem lessened):

\(^{10}\text{We exclude manager's private benefits from the social welfare function. This assumption is only for simplicity and entails no loss of generality. Alternatively we could assume that manager's utility belongs to the welfare function of the region, but still the high effort is socially optimal as in Qian and Roland (1994).}\)

\(^{11}\text{The idea that differences in objectives between governments and the private sector are at the heart of the soft budget constraint syndrome is also found in Shleifer and Vishny (1994), Segal (1994) and Qian and Roland (1994).}\)
1. some among those potentially unemployed will obtain jobs otherwise not available in their region of origin;

2. migration will reduce the congestion of the local public good provided by the region where migrants come from.

Instead, the welfare effects of immigration on region B are not clearcut and is unclear whether region B would lose or gain from immigration. In fact, on the one hand, those of the immigrants who find a job contribute to increase the total amount of profits generated in the host region. On the other hand, all of the immigrants concur to make the access to the local public good more congested. If the profit gains outweigh the enhanced congestion costs, migration will bring about a positive externality on region B; otherwise, the externality will be negative. It is apparent that the number of workers migrating from one region to the other will crucially determine the extent of the loss of region A and of the externality on region B.

Without migration, refinancing or closing the firm would exclusively be region A's own business. Yet the mere possibility of labor migration endogenously creates a link between the strategic firm and the welfare of region B. In turn, the actual shifts of labor between regions depend on mobility costs, which are then the ultimate determinants of the nature of externalities between regions.

3.3.1 The provision of local public goods

Each region provides a local congestible public good whose consumption is non-excludable within regions but perfectly excludable between regions. In other words, each resident in the region consumes the good in equal shares, while drawing diminishing satisfaction from consumption as the number of users rises.

The resources necessary to supply the good are raised by levying distortionary taxes on firms' profits, which causes a positive shadow cost of taxation equal to $\lambda$ for all the parties. To rule out unnecessary complications, we assume that workers are not taxed\textsuperscript{12}

\textsuperscript{12}We could defend this assumption assuming that workers' utility is determined by his rent: then the rent cannot be taxed because the workers would slack, à la Shapiro and Stiglitz.
The welfare function of each region is the sum of its firms’ profits and its citizens’ utilities. Region A maximizes

$$W^A = \Pi^A + \sum_i U_i^A$$

where $\Pi^A$ are aggregate profits in region A and $U_i^A$ is the utility of citizen $i$ of region A.

We postulate a very simple form of the individuals’ utility function:

$$U_i^A = w + \gamma g_A$$

where $w$ represents the rent accrued to each worker when employed and $g_A$ is per-capita consumption of the public good, i.e. $g_A = \frac{G_A}{\hat{N}_A}$, where $G_A$ is the total amount of public good offered in region A and $\hat{N}_A$ is the number of residents in region A. We assume that $G_A$ and $G_B$ are exogenously given, which is an extreme and simplifying assumption. However, our results do not hinge on the availability of a fixed supply of the public good. Our argument only requires that migration cause some congestion of the public good in the host region or a larger spending in the public good with its associated surge in distortionary taxation.

Finally, we assume a symmetric situation at the beginning: $N_A = N_B = N$, $G_A = G_B = G$ and full employment in both regions. These assumptions ensure that, at the outset, there are no incentives to migration for anybody.

### 3.3.2 Migration

We describe migration as a rational choice driven, as usual, by demand-side and supply-side factors.

Demand-side factors are summarized in the assumption that a given number $q$ of jobs is made available for immigrants in region B. Taking these jobs produces a

---

13 Obviously the difference between $N_k$ and $\hat{N}_k$ is the number of migrants.

14 The assumption that jobs are only taken by immigrants is not very restrictive in the current set-up, which rules out the existence of initial unemployment in both regions. In a more general
unitary rent \( w \) to migrants and a unitary net profit of \( \pi > 0 \) to the shareholders of region B.

Overall, our model embodies the view that migrants do not cause tension in the formal labor market of the host region, which we deem plausible both in theory and practice.

Piore’s theory of segmented labor markets\(^{15}\) predicts that migrants, for their characteristics of ‘target earners’, would be naturally inclined to take jobs in the secondary labor market. Piore also provides supportive anecdotal evidence on Mexican migrations into the US. Moreover, even without taking Piore’s point of view, available empirical evidence has been generally unable to detect anything but negligible effects of immigration on natives’ earnings and job opportunities\(^{16}\).

On the supply side, the migration decision is affected by three elements. First, migrants are attracted into region B by the uncertain prospect of getting a job\(^{17}\) at the prevailing wage \( w \). Second, migrants compare the different degrees of congestion in the use of the public good in the two regions. They take into account that migrations make the use of the public good more congested in the host region and less congested in the source region. Third, upon migrating, each migrant has to bear a mobility cost \( c > 0 \).

The presence of a cost of migration can be defended in many ways. Migration entails psychological costs (leaving friends behind is sad), costs associated to social integration in the new region (say, a new language to learn or different habits to get used to) and liquidity costs due to the inability of realizing the full value of the assets one leaves in the place of origin.

Let \( z \) be the number of immigrants. We assume

\(^{15}\)Piore (1979) views the labor market of capitalist economies as intrinsically segmented between a formal, protected and unionized, part and an informal part where menial and noxious jobs concentrate.

\(^{16}\)Borjas (1990) provides a wealth of documentation on the US experience.

\(^{17}\)That migration entails a random job lottery in the destination region is a standard feature of migration models since Harris and Todaro (1979)
**Assumption 19** \( w + \gamma \frac{G}{N+q} - c > \gamma \frac{G}{N-q} \)

Assumption 19 sets an upper bound on \( c \). An incentive to migrate is established such that slackness of migrants willing to take the available job openings in region B never arises: \( z > q \)\(^{18}\).

Under our set of assumptions, migration certainly relieves, at least partially, region A from its employment concerns, while region B may or may not be hurt by migrations. A brief discussion of two polar cases, in which region B is respectively a sure gainer and a sure loser from migration, helps further clarify this point. If discriminatory government policies effectively prevent migrants from consuming the public good\(^{19}\), then region B would enjoy a costless rise in profits, without suffering from public good congestion. Instead, region B would certainly lose if all of the job openings in region B were taken by natives, since this would result in wasteful (given the positive mobility and congestion costs) competition for a constant number of jobs in region B.

Finally, we assume that the number of migrants will be determined so that, at the margin, the last worker who migrates will be indifferent between staying in region A unemployed and migrating to region B.

By our assumption, the number of migrants \( z \) is determined by the requirement that the expected utilities of moving to region B (and being either employed or unemployed there) and staying unemployed in region A be equal:

\[
\frac{q}{z} w + \gamma \frac{G}{N+z} - c = \gamma \frac{G}{N-z} \tag{3.1}
\]

Equation 3.1 implicitly defines \( z^* = z^*(q, c, G, \gamma) \). From the implicit function theorem we derive

\[
z_q^*(q, c, G, \gamma) > 0 \]

\[
z_\gamma^*(q, c, G, \gamma) < 0
\]

\(^{18}\)We are also implicitly assuming \( \frac{q}{L} w + \gamma \frac{G}{N+L} - c < \gamma \frac{G}{N-L} \) i.e. that not all of the \( L \) unemployed workers will migrate. This represents a lower bound on \( c \)

\(^{19}\)This may, but need not, be the case for the access to some social services, such as police surveillance on property in the suburbs where immigrants usually concentrate
$z^*(q, c, G, \gamma) < 0$

and

$z^c(q, c, G, \gamma) < 0$.

These results are quite intuitive. More jobs in region B make migration more attractive; a larger provision of the public good makes migration less worthwhile; an increased preference for the public good discourages people to move away from a less congested region to a more congested one. Finally, higher mobility costs obviously reduce the level of migration.

We have now described the building blocks of our model. In the next sections, we relate mobility costs to decentralization and centralization of decision-taking and to the hardness of the budget constraint \(^{20}\).

### 3.4 Externalities and mobility costs

Now we compute the size of the loss ($l$) due to the closure of the firm in region A and the size of the externality ($b$) that closing the firm causes on region B.

In order to evaluate $l$ and $b$, we must first decide who enters the objective function of the regional governments. The problem is the following: suppose it is known at time $t$ that some members of the population will migrate inside or outside the region at time $t + 1$. Should a regional government exclusively worry about the welfare of those who are currently members of the population at $t$ or rather take into account future members as well (and, symmetrically, ignore the welfare of those who will no longer be inhabitants of the region tomorrow)? In other words, is regional welfare to be defined over citizens or over ‘those left behind’ by migrants?

In our problem, this is relevant to determine the amount of the transfer to pay to the firm in case of distress. The sum to pay obviously depends on the social loss, which in turns depends, in an utilitarian fashion, on who is considered part of the

\(^{20}\)For brevity, from here onwards, we omit reporting $G$ and $\gamma$ as arguments of the migration function.
regional community by the local government.

In the Anglo-Saxon world, non-resident nationals are usually excluded from income tax jurisdictions, in accordance with the 'schedular' tax system. In the early seventies, Bhagwati proposed the adoption of a 'global' income taxation system. The main rationale underlying Bhagwati’s proposal was equitative. ‘Taxing the brain drain’ was a means to let 'those left behind’ enjoy part of the benefits accruing to highly skilled workers migrating abroad from Less Developed Countries. If we take this approach, it is citizenship, rather than residence, which determines whose welfare migrants belong to.

While not denying the importance of equitative considerations\(^ {21} \), we find the 'schedular' system analytically more appealing. If we accept that governments are inclined to please their voters to assure themselves reelection, it seems natural to assume that only the preferences of today's (rather than future or past) voters matter in taking public policy decisions.

This implies that the loss and the externality that we compute only refers to the population present in the region at the time when the decision about the subsidy to the firm is actually taken.

It is easy to compute

\[ l(z^*(q,c)) = L_F w - N \left[ \frac{\gamma G}{N-z} - \frac{\gamma G}{N} \right] = L_F w - \gamma G \frac{z}{N-z}. \]  

\[ b(z^*(q,c)) = N \left[ \frac{\gamma G}{N-z} - \frac{\gamma G}{N+z} \right] - q \pi = \gamma G \frac{z}{N+z} - q \pi. \]  

(3.2) 

(3.3)

Notice also that

\[ l(z^*(q,c)) + b(z^*(q,c)) = L_F w - q \pi - \gamma G \frac{2z^2}{(N+z)(N-z)} \]

\(^{21}\)We refer to Bhagwati (1988) for a short survey of the main issues.
and using equation 3.1 we know that \( \frac{\partial}{\partial z} w - c = 2\gamma G_B \frac{\hat{z}}{(N+z)(N-z)} \) so that

\[
l(z^*(q, c)) + b(z^*(q, c)) = LFW - q(\pi + w) + z(q, c)c. \tag{3.4}
\]

Notice that both \( l \) and \( b \) are functions of \( z \), i.e. the stake of each region is endogenously determined by the flow of migrants, which is in turn positively affected by job creation in region B and negatively by the costs of mobility.

### 3.5 Mobility costs and the soft budget constraint

In the section above we have derived the conclusion that the nature of externalities is the crucial variable in order to assess whether centralization or decentralization implements a harder budget constraint. We must now relate this result to our migration model. First we have the following

**Lemma 14**

\[
\frac{d}{dc}[l(z^*(q, c))] > 0
\]

\[
\frac{d}{dc}[b(z^*(q, c))] < 0
\]

**Proof** \( \frac{d[l(z^*(q, c))]}{dc} \) has the same sign as \( \frac{d}{dc}[-\frac{z}{N-z}] = -\frac{z_N}{(N-z)^2} \) and since \( z_c < 0 \) the result follows.

\( \frac{d[b(z^*(q, c))]}{dc} \) has the same sign as \( \frac{d}{dc}[\frac{z}{N+z}] = \frac{z_{cN}}{(N+z)^2} \) and again the result follows from \( z_c < 0. \)

From Proposition 14 and Lemma 14 we derive

**Corollary 8** Let \( \check{c} \) be such that \( b(z^*(q, \check{c})) = 0. \) Then a necessary condition for decentralization to be socially preferable to centralization is that \( c < \check{c}. \)

This is our main result: mobility costs are a crucial factor in determining the relative performance of a centralized vs. a decentralized system. In particular, unless
mobility costs for workers are low, decentralizing entails no advantages for the society as a whole.

These findings lend themselves to policy discussion.

The cost of mobility $c$ can be thought as partly exogenous (e.g., how different are the regions in terms of language, lifestyle, habits) and partly policy-determined (through, say, bureaucratic impediments and discrimination). While we defer the analysis of the endogenous part of mobility costs later in this section, the exogenous part of the cost can be fruitfully related to the current debate on federalism in various parts of the world.

One of the most often mentioned arguments of the folklore on the topic is that federations should be composed by 'similar' regions. While the folklore argument does not specify what is really meant by 'similarity', in our model it has a precise analytical content: a low $c$. So our model can be seen as providing an analytical framework to the folklore: Lombardy is more likely to gain than Sicily from joining a 'Europe of regions'. Along the same lines of reasoning, some light can be shed on another policy issue, which is of paramount importance for African governments. Since the late eighties, after decades of dirigism, a few African governments undertook the task of encouraging the private sector by closing loss-making public and parastatal enterprises. Further progress in this process may be hampered, rather than favored, by the introduction of fiscal federalism in states ridden with ethnical diversities between regions. Since $c$ is high within African states, fiscal decentralization and delegation of responsibilities to local governments should be pursued with caution.

Our result can be (at least broadly) related to the current debate about Europe as well. In fact, the demand for fiscal federalism goes hand in hand with the process of progressive elimination of barriers to the mobility of citizens among Eu
cpean nations. The claim usually made is that fiscal federalism makes the budget constraint of the public sector harder, since each taxpayer can better monitor the destination of the taxes collected in a federal system. The enhanced monitoring ability will imply that governments will be more reluctant to start bad projects. According to our theory, a necessary condition for the commitment to hardening the budget
constraint by a decentralized government to be enforceable is that barriers to labor
mobility are lowered enough. So our theory finds a link between the reduction of
cost in labor mobility and the diffusion of powers at the local level. Wellisch (1994)
reaches a comparable conclusion, although his discussion is unrelated to the soft
budget constraint problem we have been discussing here. His main point is that,
under perfect household mobility, the decentralized provision of pure public goods
generating spillovers across communities is socially efficient.

As we mentioned above, while part of the determinants of $c$ is exogenous, some
others are influenced by policy decisions. Now we investigate the impact of a change
in $c$, the mobility cost, on the hardness of the budget constraint in the centralized
and in the decentralized setting. In other words, we ask the following questions: is
reducing mobility costs always a good thing? Is it more valuable under centralization
or under decentralization?

First we have

**Lemma 15** $\frac{d}{dc}(l(z(c)) + b(z(c))) > 0$

**Proof**

By equation 3.4, all we need to compute is $\frac{d}{dc}[z(c)c]$. Using equation 3.1 we find
that $\frac{d}{dc}[z(c)c] = -\frac{2ze^{N}}{(N-z^{2})^{2}} > 0$ since $z _{c} < 0$.

By Lemma 15 and noting that, when a soft budget constraint is implemented, the
total payoff under centralization is $-(l + b) + \frac{(l+b)^{2}}{2c(1+c)}$ we have

**Proposition 15** A (costless) reduction in $c$ is always welfare-improving in a central-
ized setting.

**Proof** Notice first that a reduction of $c$ reduces $l(c) + b(c)$ so it has beneficial effect
in hardening the budget constraint. What about the case of soft budget constraint?
In centralization, total payoff is $-(l + b) + \frac{(l+b)^{2}}{2c(1+c)}$ whose derivative with respect to $c$
is clearly positive.

This result implies that easing labor mobility is always a goal to be achieved in
centralization.
A similar result holds in decentralization only if \( b < 0 \), i.e. \( c > \hat{c} \). No general conclusion can be derived for the decentralized case when \( b > 0 \) i.e. \( c < \hat{c} \).

We have also the following

**Proposition 16** Suppose \( l > b > l - \rho(1+\lambda) \) and assume that a soft budget constraint is implemented both in centralization and decentralization. Then the marginal gain of a reduction in \( c \) is higher in decentralization if \( c > \hat{c} \). The opposite holds if \( c \leq \hat{c} \).

**Proof** The payoffs under centralization and decentralization coincide for \( c = \hat{c} \). Their difference is \( \frac{b^2}{2\rho(1+\lambda)} \). The derivative of the difference with respect to \( c \) is \( \frac{2bbc}{2\rho(1+\lambda)} \). The derivative is positive if \( b < 0 \) i.e. \( c > \hat{c} \) and negative if \( b > 0 \) i.e. \( c < \hat{c} \). We know from the Proposition above that the derivative with respect to \( c \) of total payoff under centralization is always positive. This means that a reduction of \( c \) reduces the distance between the two payoffs for \( c > \hat{c} \) (hence it is more valuable under decentralization) and increase the distance for \( c < \hat{c} \) (hence it is more valuable under centralization).

\[ \square \]

The Proposition above implies that it is not possible to rank the effort in reducing mobility costs between a centralized and a decentralized setting. It all depends on the initial value of \( c \).

### 3.6 Many regions

In this section we explore how the presence of many regions which are potential destinations of migrants from region A changes our results. The main result of this section is that many regions of destination reinforce the basic results proved above, i.e. when externalities among regions are positive, decentralization implements a harder budget constraint and vice versa. To simplify the exposition we focus on the case where there are two potential host regions. Suppose total externality \( b \) is given and assume that each regions has an externality \(-b/2\) from the firm when closed.

As before, we must study two cases:

1. \( b > 0 \)
b) $b < 0$

In case a) let $\{s_B, t_B\}$ and $\{s_C, t_C\}$ be the contracts offered by region B and C respectively to region A in order to keep the firm open. We impose again the limited liability assumption.

Region A solves

$$\max_E \ E(1 + \lambda)(s_B + s_C) - (1 - E)[l - (1 + \lambda)(t_B + t_C)] - \rho(1 + \lambda)E^2/2.$$ 

The solution is

$$E = \frac{l + (1 + \lambda)(s_B + s_C - t_B - t_C)}{\rho(1 + \lambda)}.$$ 

Region B problem is

$$\max_E \ [-E(1 + \lambda)s_B - (1 - E)[b/2 + (1 + \lambda)t_B]]$$

subject to $s_B, t_B \geq 0$. The solution of this problem is $t_B = 0$ and $s_B = 0$ if $b > 2l$ and

$$s_B = \max\{0, \frac{b - 2l}{4(1 + \lambda)} - \frac{s_C}{2}\}$$

otherwise. Notice that now the range in which transfers do not occur is $b < 2l$, while before it was $b < l$. The presence of multiple regions makes transfers more unlikely.

A symmetric conclusion can be reached for region C. In a symmetric Nash equilibrium when transfers occur it holds

$$s_B = s_C = \frac{b - 2l}{6(1 + \lambda)}$$

so that

$$s_B + s_C = \frac{b - 2l}{3(1 + \lambda)}.$$
Notice that for $b > 0$ it always holds

$$\frac{b - 2l}{3(1 + \lambda)} < \frac{b - l}{2(1 + \lambda)}$$

so that $E$ is now lower than the expression we found for only one region.

We can generalize the discussion above in

**Proposition 17** If $b > 0$, the higher the number of regions over which the externality is divided, the harder the budget constraint.

In case b), $b < 0$ and region A maximizes:

$$E(1 + \lambda)(s_B + s_C) - (1 - E)[l - (1 + \lambda)(t_B + t_C)] - \rho(1 + \lambda)E^2/2$$

and the solution is

$$E = \frac{l - (1 + \lambda)(t_B + t_C - s_B - s_C)}{\rho(1 + \lambda)}.$$

Region B solves instead

$$\max\quad -E(1 + \lambda)s_B - (1 - E)[b + (1 + \lambda)t_B].$$

subject to $s_B, t_B \geq 0$. The solution of this problem is $s_B = 0$ and $t_B = 0$ if $b > 2[l - \rho(1 + \lambda)]$ and

$$t_B = \max\{0, \frac{l - b/2}{2(1 + \lambda)} - \frac{\rho + t_C}{2}\}$$

for $b < 2[l - \rho(1 + \lambda)]$.

Notice again that the range in which no transfers occur is $b < 2[l - \rho(1 + \lambda)]$, twice as large as before.

In a symmetric equilibrium when transfers occurs it holds

$$t_B = t_C = \frac{l - b/2}{3(1 + \lambda)} - \frac{\rho}{3}.$$
so that total transfers are now

\[ t_B + t_C = \frac{2l - b}{3(1 + \lambda)} - \frac{2 \rho}{3}. \]

Comparing this expression with the one found in the case of only one region we see that, when \( l + b < \rho(1 + \lambda) \) it always holds

\[ \frac{2l - b - 2 \rho(1 + \lambda)}{3(1 + \lambda)} < \frac{l - b - \rho(1 + \lambda)}{2(1 + \lambda)} \]

i.e. total transfers are lower than in the previous case. Hence \( E \) will be higher than the case where only one region was affected by the externality. We can generalize the discussion above in

**Proposition 18** If \( b < 0 \), the higher the number of regions over which the externality is divided, the softer the budget constraint.

The results above are in line with the findings of Burkart et al. (1994). They argue that in order to reduce ex-post monitoring in a managerial firm on the part of shareholders, a high dispersion of claims can be beneficial \(^{22}\). The reason is that dispersion of claims induces a free-rider problem among the shareholders of the firm who internalize the full cost of monitoring but only part of the benefits. The same happens in our model: regions B and C reduce the amount of transfers to region A because they do not fully internalize the benefits of their transfers.

### 3.7 Size of fiscal programs in centralization and decentralization

Persson and Tabellini (1994) address an important question concerning fiscal federalism: are fiscal programs larger or smaller if decentralized to the local level of a prospective federation? Their answer is roughly the following:

\(^{22}\)This corresponds to the case \( b > 0 \).
1. for fiscal programs whose benefits are broadly spread in the population, the size of the budget of a central government will be lower;

2. for fiscal programs whose benefits are localized, the size of government budget can be reduced by decentralization.

While their models are different from ours, which makes comparisons difficult, nonetheless we believe that our results have some bearing on theirs. In our model, $E$ can be thought as measuring the size of the fiscal program and benefits always get spread over both regions\(^2\), it is tempting to parallel our conclusions to theirs in the case of broadly spread benefits. We do not have a clearcut answer: $E$ is larger in a decentralized setting if $c > \hat{c}$ and smaller if $c \leq \hat{c}$. As we have already stressed, in the case where refinancing bad projects causes a negative externality on region B, decentralization will imply an inefficiently large amount of resources spent in the bail-out, while centralization correctly internalizes the negative externality, reducing the size of the fiscal program.

### 3.8 Concluding remarks

In this paper we reached the following main conclusions.

First, the argument usually found in the literature claiming that decentralization is more efficient than centralization in implementing a hard budget constraint - an example of the subsidiarity principle - depends on the assumption that positive externalities are involved among different parties in the refinancing decision. The opposite result holds in case of negative externalities.

Second, labor mobility is a means of achieving a situation where multiple regions are involved in the refinancing decision. The stake of each region will depend upon mobility costs so that the effectiveness of decentralization in implementing a hard budget constraint is enhanced by low mobility costs. This amounts to say that fiscal decentralization should be preceded (or go together) by policies favoring mobility.

Third, given the total amount of externality, the number of parties affected is a

\(^2\)More precisely, in our model the stake of each region depends on $c$. 

110
crucial ingredient in determining the hardness of budget constraint.

Fourth, no general conclusion can be drawn on the size of local vs. central fiscal programs. In other words, it is not clear a priori that decentralization is a way of reducing government intervention.

We have modelled the differences between centralization and decentralization in terms of differences in the information available to the political authority. However, while our approach is consistent with similar models in the literature, we still do not have a complete theory of how information is collected and how it spreads inside organizations. The analysis of these problems will hopefully be the object of future research.
Chapter 4

References


Admati, A., P. Pfleiderer, and J. Zechner (1994), Large shareholders activism, risk sharing, and financial market equilibrium, *Journal of Political Economy*


Bebchuk, L. and R. Picker (1994), Bankruptcy rules, managerial entrenchment, and firm-specific human capital, mimeo, University of Chicago

Berglöf, E. and E.L. von Thadden (1994), Capital structure with multiple investors,


Bolton, P. and D. S. Scharfstein (1993), Optimal debt structure with multiple creditors, WP CEPR, No. 32.

Borjas, G.J. (1990), Friends or Strangers: the Impact of Immigrants on the U.S. Economy, New York: Basic Books


Boycko, M., A. Shleifer and R. Vishny (1993), Privatizing Russia, Harvard University, mimeo


Hart, O.D. (1993), An economist’s view of fiduciary duty, WP FMG


Herman, E.S. (1981), *Corporate control, corporate power*, Cambridge University Press.
Hermalin, B. and M.S. Weisbach (1988), The determinants of board composition, 

Holmström, B. (1982), Managerial incentive problems - a dynamic perspective, mimeo
Northwestern University, Evanston.

Holmström, B. and J. Tirole (1993), Market liquidity and performance monitoring, 
*Journal of Political Economy*, 101:

Hoshi, T, A. Kashyap and D. Scharfstein (1991), The role of banks in reducing the 
costs of financial distress in Japan, *Journal of Financial Economics*, 20:

Ihamuotila, M. (1993), Ownership structure and investor-diversification: The choice 
of investment and financing, mimeo, Yale.

Kahan, M. and B. Tuckman (1993), Private vs. public lending: Evidence from 
covenants, mimeo.

Kaplan, S.N. (1994), Top executives, turnover and firm performance in Germany, WP 
CEPR, No. 45.

Klibanoff, P. and J. Morduch (1993), Decentralization, externalities and efficiency, 
mimeo, Northwestern University

61:629-656.

Kornai, J. (1979), *Economics of Shortage*, Amsterdam: North Holland

Lummer, S.L. and J.J. McConnell (1989), Further Evidence on the Bank Lending 
Process and the Capital-Market Response to Bank Loan Agreements, *Journal 

Maskin, E. (1994), Theories of the soft budget constraint, mimeo, Harvard University

Miles, D. (1992), Testing for short-termism in the UK stock market, WP Bank of 
England.

Morck, R., R. Shleifer, and Vishny, A. (1988), Management ownership and market 


Povel, P. (1994), Multiple banking, mimeo, LSE

Qian, Y. and G. Roland (1994), Regional Decentralization and the Soft Budget Constraint: the Case of China, mimeo, Ecare, Brussels


Rajan, R. and A. Winton (1994), Covenants and collateral as incentives to monitor, WP University of Chicago.


Schmidt, K. (1991), The costs and benefits of privatization, mimeo, MIT.

Segal, I. (1994), Monopoly and soft budget constraint, mimeo, Harvard University


Winton, A. (1992), Limitation of liability and the ownership structure of the firm, WP Kellogg School, Northwestern University.


Zingales, L. (1994), Insider Ownership and the decision to go public, mimeo, University of Chicago.