Are banks excessively monitored?

Urs W. Birchler*

Nov 22, 2000

Abstract

Insufficient monitoring by depositors, and thus a lack of market discipline, are often seen as a typical feature of banks. We show that the opposite may be the case. Banks, defined as firms that borrow from a large number of partially uninformed investors, have a tendency to be excessively monitored by informed investors. This is shown in a model of intermediation in which heterogenous investors choose whether they want to monitor the intermediary or not. We also find that bank finance is preferable to non-bank finance when assets are relatively safe or opaque. The model which is set in a banking context may be applicable to a wider range of information problems.

*I thank Jürg Blum, Monika Bütler, Bhagwan Chowdhry, Harris Dellas, Dominik Egli, Andréa Maechler, and Valter Sorana for helpful comments. The paper was presented at the 2000 CEPR European Summer Symposium in Financial Markets (ESSFM), Gerzensee, as well as in research seminars at CentER Tilburg, and the University of Bern. Errors are my own. The paper does not reflect the views of Swiss National Bank. Address: Swiss National Bank, P.O. Box, CH-8022 Zurich, Switzerland; e-mail: urs.birchler@snb.ch

1 Introduction

The stability of banks hinges on adequate supervision, either by the market, or by a special authority. A failure of the market, i.e. of potential or actual depositors, to monitor banks may create a need for bank regulation (see Dewatripont and Tirole, 1994). The dominant view among economists is that banks are indeed *insufficiently* monitored. It is argued that banks are typically financed by a large number of relatively small and unsophisticated depositors. To such depositors, the costs of monitoring a bank are likely to exceed any possible (private) returns. Prominent advocates of this view are Dewatripont and Tirole (1994):

> "Bank debt is primarily held by small depositors. Such depositors are most often unsophisticated, in that they are unable to understand the intricacies of balance and off-balance sheet activities. More fundamentally the thousands or hundreds of thousands of customers of a bank have little individual incentive to perform the various monitoring functions. This free-riding gives rise to a need for private or public representatives of depositors. We call this the *representation hypothesis*." (p. 31f., their emphasis)

Similarly Freixas and Rochet (1997) state that:

> "Contrarily to nonfinancial firms, the debt of which is held in majority by 'professional investors' (i.e. banks, venture capitalists, or 'informed' private investors), the debt of banks (and insurance companies) is held in large part by uninformed, dispersed small agents (mostly households) who are not in a position to monitor banks' activities." (p. 264)

Lack of market discipline is thus a characteristic of banking *per se*, and it explains why banks need supervision.[1] Measures like deposit insurance, despite their merits,[2] may be dangerous, as they may further undermine market discipline (see, e.g., Greenbaum and Thakor, 1995; Freixas and Rochet, 1997).

[1] According to Dewatripont and Tirole this is the only robust argument in favour of banking regulation.

[2] Most importantly a lower probability of bank runs, see Diamond and Dybvig (1983).

The prevailing "monitoring deficit" view highlights important aspects of financial intermediation like the public good nature of information and the danger of moral hazard. Nevertheless, to our view, it is incomplete, as it overlooks other important aspects of financial intermediation. One such aspect is the influence of the intermediary's choice of deposit contracts on investors' demand for information. In the present article we try to complement the common view, by analyzing *endogenous* monitoring of a financial intermediary by investors.[3] Making the demand for information endogenous to contracts has interesting consequences. We find a result that is at odds with the common view of insufficient monitoring: In our model banks are *excessively* monitored.

Our line of argument can be summarized as follows. The main actor is a financial intermediary. This intermediary can either be a bank, borrowing from uninformed investors (see Dewatripont and Tirole, 1995) or a non-bank, only borrowing from informed investors. Through the intermediary investors can finance a risky asset. There is only one type of risky asset; unlike in Diamond (1984), the intermediary cannot diversify risk.[4] Investors thus have an interest to collect information on the intermediary's earnings prospects, i.e. to monitor. In the model, they can acquire a costly private signal on the future returns of the intermediary's assets. We use a model in the tradition of Cukierman (1980) where investors can update prior expectations by receiving a signal.[5] We find that under *bank* intermediation the demand for the signal is higher than in the benchmark case of a hypothetical unintermediated economy in which investors can directly hold the risky asset. This is

[3] We are thus in line with Allen (2000) who identifies as one research priority "to incorporate information in a fully endogeneous way into microeconomic theory" (p. 148). A recent example for the importance of endogenous monitoring cost is given in Xu (2000).

[4] The benefits of intermediation are thus given and limited. But even if perfect diversification were possible, as in Diamond (1984), it might not be optimal for the intermediary. Intermediaries do not fully diversify in the presence of a fixed monitoring cost per project or per entrepreneur financed (Krasa and Villamil, 1992), particularly when they can choose project size (Hellwig, 1998).

[5] In Cukierman (1980), the signal improves the precision of return expectations; in our model it primarily affects the level of expected returns.

In both models, information gathering occurs before contracts are made. Another family of models builds upon post-contractual information acquisition. Such models normally focus on the liquidation decision (see e.g. Calomiris and Kahn, 1991) and on renegotiation possibilities (see e.g. Park 2000).

what we call excess monitoring.[6]

While this result contrasts with the common view, our model also predicts some features of financial intermediation that seem more standard. We find that relatively safe or informationally opaque assets tend to be financed by banks. Further, banks issue debt-like instruments while non-bank intermediaries issue instruments that are a mix between debt and equity.

We also find a result that may have a wider range of applications. To present our argument, we distinguish between projects that are only worthwhile if some good news arrives (called entry-games) and projects that are worthwhile unless bad news arrives (exit-games). A bank deposit is an exit-game (as it is attractive to uninformed investors), while lending to a non-bank is an entry-game. Analyzing such games we find that the value of information *in*creases in potential returns in entry-games and *de*creases in exit-games. It is highest in games that a priori just about break even. This may e.g. explain why investors do not seem to care about the probability of a currency crisis when a country is either in very poor or in extremely good shape, but much so when it is in an intermediate condition.

The article is structured as follows: In section 2 we propose a simple example providing some intuition for the main result of this article. In section 3 we model a non-intermediated economy as a benchmark against which we will measure the outcome from an intermediated economy to be analyzed in section 4. In section 5 we compare monitoring and investment levels in the two economies. Section 6 concludes.

2 Entry versus exit: an example

The following example is meant to illustrate the functioning of our model economy. Neil considers going on a camping trip.[7] As most campers, he dislikes rain. He values a dry trip at one dollar and a wet trip at minus one dollar. Staying home has a value of zero. Sunshine and rain have probability 0.5 each. Neil is thus indifferent between going and staying.[8] Fortunately, before he leaves, he can call a weather service to get a forecast. The voice on the service tape will say “good” or “bad”. “Good” means that chances of rain are revised down to 0.25; “bad” means they are up to 0.75. For

[6] Under non-bank intermediation monitoring can be insufficient or excessive.

[7] The choice of a camping trip as an example is an hommage to Wallace (1988).

[8] We assume that when indifferent he takes the trip.

consistency with prior and posterior weather probabilities, both forecasts must have probability 0.5. Neil's situation is illustrated by the decision tree in Figure 1.

[Figure 1 about here]

How much would risk neutral Neil be ready to pay for the forecast? The difference in utility between being informed and uninformed is easy to compute. Informed Neil goes camping after hearing "good". (This occurs with probability 0.5 and has expected utility of half a dollar.) He stays home after hearing "bad" (with utility zero). Uninformed he is indifferent (both alternatives having expected utility zero). The forecast (utility with forecast minus utility without) is thus worth one quarter.[9]

To make Neil's problem a bit more interesting, let us assume that his utility from a trip is slightly modified. Beer on the campground may be less or more expensive than normal. For simplicity he always has a drink on a dry trip, but never on a wet trip. His utility from a dry trip is now $1 + \epsilon$, with $\epsilon \gtrless 0$ (beer cheap or expensive). We assume that ϵ, in absolute terms, is sufficiently small not to make the weather forecast irrelevant.[10] Neil knows ϵ in advance.

Informed Neil still goes for the trip after a good forecast and stays home after a bad one. Uninformed Neil will go for the trip when $\epsilon \geq 0$ and stay home when $\epsilon < 0$. We could also say that, if $\epsilon < 0$, Neil decides not to make the trip, but would want to enter late in case of a good weather forecast. Conversely, if $\epsilon \geq 0$, he would enroll for the trip but would want to exit in case of a bad weather forecast. We will thus formulate two definitions for later use.

Definition 1 *The game represented in Figure 1 is*

- *an* entry-*game if* $\epsilon < 0$*,*
- *an* exit-*game if* $\epsilon \geq 0$*.*

We found the value of the weather forecast (when beer has no influence, i.e. when $\epsilon = 0$) to be one quarter. How does the price of beer influence this

[9] In numbers: $\frac{1}{2}\left(\frac{3}{4}1 + \frac{1}{4}(-1)\right) + \frac{1}{2}0 - 0 = \frac{1}{4}$.

[10] It is sufficient that $-\frac{2}{3} < \epsilon < 2$ to exclude the two trivial cases in which Neil ignores the weather forecast.

value? The value of the weather forecast (the value of the option to enter late or exit early) now depends on ϵ, i.e. on whether the trip is an entry- or an exit-game. We denote the values of entry by V_{ϵ^-} and of exit by V_{ϵ^+}, respectively. It can easily be derived from Figure 1 that

$$V_{\epsilon^-} = \frac{1}{4} + \frac{3}{8}\epsilon, \text{ where } \epsilon < 0, \quad (1)$$

$$V_{\epsilon^+} = \frac{1}{4} - \frac{1}{8}\epsilon, \text{ where } \epsilon \geq 0. \quad (2)$$

Note that both values, V_{ϵ^-} and V_{ϵ^+}, are less than one quarter, as the last terms in (1) and (2) are both negative. This means that the value of information has its maximum at $\epsilon = 0$, where uninformed Neil would be indifferent.

How does V react to a small increase in the price of beer, i.e. to a *decrease* in ϵ? Taking (negative) derivatives of (1) and (2) yields:

$$\frac{\partial V_{\epsilon^-}}{-\partial \epsilon} = -\frac{3}{8} < 0, \quad (3)$$

$$\frac{\partial V_{\epsilon^+}}{-\partial \epsilon} = +\frac{1}{8} > 0. \quad (4)$$

A deduction from an agent's return (here: a higher price of beer) thus *decreases* the value of information in an entry-game (the entry option becomes less valuable) and *increases* it in an exit-game (exit becomes more valuable). This rule is important in what follows, as we will apply it to bank deposits, rather than to camping trips. A bank deposit is an exit-game (it is attractive to uninformed agents). As (4) shows, the value of information on whether the depositor will get a dry or a wet trip, so to say, increases when the promised return on deposits decreases. This inverse relationship between the face value of a bank deposit and the value of information to depositors will provide the basis for the main result of this paper.

3 The non-intermediated economy

In the following sections we present our argument more formally. First, we develop the model and analyze a non-intermediated economy to serve as a benchmark for the intermediated case.

3.1 The model

The economy has two investment vehicles. The first is a (perfectly divisible) risky asset with a publicly observable per dollar return of $\tilde{y} \in \{Y, 0\}$. Second, there is a (perfectly divisible) safe asset that per dollar yields R $(0 < R < Y)$.

The risky asset's prior probabilities of "success" (Y) and of "failure" (0) are $p > 0.5$ and $(1-p)$, respectively. However, before money has to be invested, a (costly) signal $\sigma \in \{g, b\}$ becomes available. The signal, which is the same for all agents who get it, updates the odds of success and of failure. After a good signal, g, the chances of success are $q > p$ and those of failure $(1-q) < (1-p)$. After a bad signal, b, the odds are reversed.[11] The probabilities of receiving the signal g or b, henceforth u and $(1-u)$, follow from the prior and posterior probabilities p and q and are[12]

$$u = \frac{q+p-1}{2q-1}, \text{ and } 1-u = \frac{q-p}{2q-1}. \tag{5}$$

There is a continuum of investors, who are completely identified by their individual "type", i.e. by their individual cost s of observing the signal. The (deadweight) signal expenditure s is made in non-financial form (e.g. as an effort). Types are unobservable. They are uniformly distributed over the interval $s \in \left[0, \overline{S}\right]$.[13] Investors are risk-neutral and in the aggregate they have one dollar to invest.[14]

We identify the level of monitoring with the fraction k $(0 < k < 1)$ of investors who decide to acquire the signal. A fraction $1-k$ decide to get no signal; we will call their state of information n. Investors cannot communicate; their state of information (g, n, b) as well as their investment

[11] The signal tells the investor from which of two lotteries $\tilde{y}$ is drawn. The two lotteries correspond to the camping trip with a good, and a bad weather forecast, respectively in the example given above. The probabilities q and $(1-q)$ can be interpreted as the probabilities that the signal, as a forecast of success or failure, will be right, or wrong, respectively. A signal with the same structure is used in Dow & Rossiensky (1998).

[12] The solutions for u and $1-u$ in the text follow from the condition for consistency $p = uq + (1-u)(1-q)$.

[13] Individual differences in s can have several reasons in practice. The per-dollar cost of monitoring may fall with the size of an investment, with its duration, and with the sohistication of the investor, see Jackson and Kronman (1979).

[14] In Birchler (2000) a benchmark case of this model is used in which the project's expected return after a bad signal is equal to R, but where failure pay-off may be positive, allowing separating equilibria characterized by dual class debt.

Region	Y	g	n	b	Game type
O	$Y < R/q$	no	no	no	
A	$R/q \leq Y < R/p$	yes	no	no	entry-game
B	$R/p \leq Y < R/(1-q)$	yes	yes	no	exit-game
C	$R/(1-q) \leq Y$	yes	yes	yes	

Table 1: Investment into risky asset as a function of potential return and the signal received

decisions are strictly private knowledge.[15] Monitoring in this model thus is the (pre-contractual) acquisition of information on returns from a particular asset ("screening"), rather than on the (post-contractual) behavior of an intermediary. This restriction simplifies the comparison of monitoring levels with and without an intermediary.

The time-line of events is the following: In $t = 0$ investors learn about model parameters including their (privately known) s. In $t = 1$ investors first have to decide about getting the signal before investing either in the risky asset or in the safe asset. In $t = 2$ the risky asset matures and its return becomes known.

3.2 Investment

We solve the game backwards. Investors first determine how to invest in each state of information, then on getting the signal. The optimal investment strategy is represented in Table 1. An investor prefers the risky to the safe asset if it has a higher (or at least the same) expected return, the expectation being conditional on his state of knowledge. The decision to invest is a function of the potential return of the risky asset, Y, and of the investor's three possible states of information. Depending on the state of knowledge, the probability of success, i.e. of Y, can be q (signal g), p (no signal), and $1-q$ (signal b). Table 1 shows in terms of these parameters, when an investor should invest in the risky asset ("yes") or buy the safe asset ("no").

The table shows that, when Y is in region O, a rational investor never invests in the risky asset. With Y in region C, he always does. In both cases

[15] This assumption is made to isolate the model from any external effects of individual information. It is less unrealistic than it may seem, as in practice information is not only costly to transfer, but also not necessarily credible.

he disregards any signal. We will not deal with these trivial cases. The two interesting cases are the ones with Y in either region A or B. In region A the investor prefers the risky asset only after a good signal; the risky asset is an entry-game in the sense of Definition 1, above. In region B the investor prefers the risky asset unless he has bad news; here the risky asset is an exit-game.

3.3 Monitoring

A rational investor will buy the signal if its value is bigger than or equal to its cost. If the value of the signal to investors is V, all investors with information cost $s \leq V$ buy the signal. Therefore, the fraction of monitoring investors is

$$k = V/\overline{S}. \tag{6}$$

To solve k we have to know the value of the signal, V. This value depends on the investment strategy summarized in Table 1. In region A the signal is only relevant if good. Conversely, in region B the signal is only relevant if it is bad. In both cases, its value is equal to the probability that the signal is relevant (u in region A, and $(1-u)$ in region B), multiplied by the net expected return from acting according to the signal (getting qY instead of R in region A, and R instead of $(1-q)\,Y$ in region B). The values of the signal in the two regions, denoted by V_A and V_B, are thus

$$V_A = u\left[qY - R\right], \tag{7}$$

$$V_B = (1-u)\left[R - (1-q)\,Y\right], \tag{8}$$

the difference in the $V's$ being due to a different investment behavior of uninformed investors. Note that both, V_A and V_B, decrease in p and increase in q. In other words, uncertainty (lower p) makes the signal more valuable, and so does a higher information content (higher q).[16]

Figure 2 illustrates the fraction of informed investors in the non-intermediated equilibrium, k, as a function of Y. The function $k(Y)$ is roof-shaped: The value of information, V, rises with Y in region A (where the risky asset is an entry-game) and falls with Y in region B (where the risky asset is an exit-game). The value of information, V, and thus the fraction of monitors, k, are highest around $Y = R/p$: For low and high values of Y

[16] This is easy to verify by use of $\frac{\partial u}{\partial p} = \frac{1}{2q-1}$ and $\frac{\partial u}{\partial q} = -\frac{2p-1}{(2q-1)^2} = \frac{2u-1}{2q-1}$.

($Y = R/q + \alpha$ or $Y = R/(1-q) - \alpha$, with small α), investment mistakes are not very expensive. At $Y = R/q + \alpha$ an uninformed investor misses a small opportunity by not investing, although the signal would have been good. At $Y = R/(1-q) - \alpha$ the expected loss from blindly investing, although the signal is bad, is small as well. In both cases, the signal can be ignored without severe consequences. In contrast, in the middle range of $\overline{Y}$ (around $\overline{Y} = R/p$) ignorance is most expensive. Here, the expected return foregone by a wrong investment decision is relatively large. As can be seen in Figure 2, at $Y = R/p$, the fractions of monitoring investors, $k_A = V_A/\overline{S}$ and $k_B = V_B/\overline{S}$, take the same value. The value of the signal at $Y = R/p$ is

$$V_A = u\frac{(q-p)}{p}R \equiv (1-u)\,\frac{(q+p-1)}{p}R = V_B. \tag{9}$$

We have assumed that $0 < k < 1$, i.e. that neither the set of informed nor the set of uninformed investors are empty. Figure 2 makes clear what the requirements are. For $k > 0$, Y must satisfy $R/p < Y < R/(1-q)$ (as $\min s = 0$). For $k < 1$, the peak of k at R/p must be below unity. According to (9), this requires that

$$\overline{S} > u\frac{q-p}{p}R. \tag{10}$$

For what follows we assume that (10) holds.

The roof-shape of $k(Y)$ or of $V(Y)$ may explain information demand in a much wider range of situations. In the labor market, e.g., information on a job opportunity (or on a candidate) is most valuable when the decision-maker is roughly indifferent about accepting or declining. It is not very valuable when an offer looks either poor or irresistible.[17]

We will use the non-intermediated economy as a benchmark for monitoring levels under intermediation. A monitoring level that exceeds or falls short of the corresponding monitoring level in the non-intermediated economy as depicted in Figure 2 will be called excessive or deficient.

[Figure 2 about here]

[17] This example has been suggested by Andréa Maechler.

4 Intermediation

4.1 The intermediary and contracts

We now examine an economy with an intermediary. In such an economy investors do not have direct access to the risky asset; they can only lend to the intermediary or invest in the safe asset. The intermediary is a monopolist and owns the risky asset.[18] She has no influence on its parameters, though. Her only role is to make the risky asset accessible to investors.

The intermediary, like investors, is risk-neutral. She has no wealth of her own, but can finance the risky asset by borrowing from investors. To do so, she must offer contracts which she does on "take-it-or-leave-it" terms. We assume that the intermediary has to announce contracts in period $t = 0$, i.e. before investors take any decisions as to the signal and to investment. In $t = 1$, after having learned about available contracts, investors decide about getting the signal, i.e. becoming monitors. The intermediary gets the signal for free. Still in $t = 1$, investors can accept a contract or buy the safe asset. We will call an investor who accepts a contract a "depositor".[19] The number of depositors is not publicly observable, nor is the intermediary's investment decision. In fact, the intermediary invests all the funds she gets in the risky asset.[20] In $t = 2$ she receives the return from the risky asset and shares it with depositors according to contracts.

The choice of contracts in this simple economy is very limited. Contractual payments can only depend on verifiable states of nature. The only such states are success and failure of the risky asset. The signal, the state of knowledge of individual investors or their lending decisions have been assumed to be private knowledge.[21] Furthermore, limited liability prevents contractual payments in case of failure. Therefore, contracts can only have one argument, the promised payment in case of success. We will denote this payment by D.

In the presence of an intermediary there are three decisions to be taken: First, the intermediary offers contracts that maximize her profits. Then,

[18] The assumption of a monopoly simplifies the analysis. For our qualitative results, some market power of the intermediary would suffice.

[19] Investors who are indifferent are assumed to choose the alternative preferred by the intermediary.

[20] We will prove this in the Appendix.

[21] We have assumed that investment decisions are not observable to investors; contracts therefore cannot be written on the fraction of investors who become depositors.

D	g	n	b	Contract type	Game type
$R/q \leq D < R/p$	yes	no	no	monitoring contract	entry-game
$R/p \leq D < R/(1-q)$	yes	yes	no	banking contract	exit-game

Table 2: The depositing decision as a function of the promisedpayment and the signal received

investors decide about the signal. Finally, they choose between accepting a contract (making a deposit) and buying the safe asset. Again, optimal decisions are found by backwards induction, i.e. in the order: depositing, monitoring, and choice of optimal contracts.

4.2 Depositing

In the presence of an intermediary, investors do not decide about investment, but only about depositing. The investment decision (risky versus safe asset) is taken by the intermediary. Due to her limited liability, she always puts all the funds she manages to raise into the risky asset, even after a bad signal.[22] The investors' problem is thus very similar to the one in the non-intermediated economy. The only difference is that the return to investors when the risky asset succeeds is now D instead of Y. The optimal depositing strategy is summarized in Table 2.[23] The two relevant regions differ with respect to the behavior of uninformed investors. We call a contract with $R/q \leq D < R/p$ a "monitoring contract". Such a contract is only attractive to successful monitors, i.e. to investors with a good signal (who attach probability q to success). The monitoring contract, in our terminology, is an entry-game.

In contrast, a contract with $R/p \leq D < R/(1-q)$ is called a "banking contract". It is not only attractive to investors with good news, but also to uninformed investors (to whom success has probability p). A banking contract is an exit-game. The interpretation of a bank as a firm that borrows from uninformed depositors follows Dewatripont and Tirole (1994, p. 32).

[22] The intermediary cannot commit herself to an investment strategy. After a bad signal she thus either gets no deposits (if uninformed investors do not deposit). Or, she has to pay depositors an expected income of at least R (to attract uninformed depositors). Therefore her profit from buying the safe asset after a bad signal would be zero. In contrast, betting on the risky asset yields positive profit with a probability of $(1-q)$.

[23] We ignore the two trivial cases in which investors, irrespective of their state of knowledge, never make a deposit ($D < R/q$) or where they always do ($R/(1-q) \leq D$).

4.3 Monitoring

The value of the signal follow from investment strategies. By analogy to signal values in regions A and B in the non-intermediated economy, signal values under a monitoring contract and under a banking contract, V_m and V_b, are

$$V_m = u\left[qD - R\right], \tag{11}$$
$$V_b = (1-u)\left[R - (1-q)\,D\right]. \tag{12}$$

If the monitoring contract were always offered in region A and the banking contract in region B, comparison of monitoring levels would be straightforward. As $D \leq Y$, $V_m \leq V_A$ and $V_b \geq V_B$, there would be too little monitoring under the monitoring contract and too much under the banking contract. Yet, this is not the case. As we will show below, the simple rule "monitoring contract in region A, banking contract in region B" does not hold. Before judging monitoring levels we thus first have to find which contract is offered for the different admissible values of Y.

4.4 The optimal contract

Under decentralized decision making aggregate revenue is maximized in the non-intermediated equilibrium discussed above in Section 3. The intermediary could implement this equilibrium by offering investors the contract $D = Y$.[24] Yet, this would wipe out, rather than maximize her profit. The (privately) optimal contract maximizes the intermediary's expected profit, under the assumption that investors follow rational information and investment strategies. It is found in two steps. First, we derive both, the optimal monitoring contract and the optimal banking contract, separately. Then, we compare profits from each contract to see which contract is offered under what circumstances.

4.4.1 The optimal monitoring contract

Under a monitoring contract, the intermediary has $k_m\,(D_m)$ depositors after a good signal (probability u), and no depositors after a bad signal. The fraction

[24] This would be the equilibrium if all bargaining power were with investors, rather than with the intermediary.

of monitoring investors, k_m, is an *in*creasing function of D_m as the monitoring contract is an entry-game. After a good signal, success has probability q. The intermediary thus solves

$$\max_{D_m} \Pi_m = k_m (D_m) uq [Y - D_m] \tag{13}$$

under the following constraints:

$$\begin{aligned} k_m = \frac{V_m}{\overline{S}} &= \frac{1}{\overline{S}} u [qD_m - R], & (14)\\ qD_m - R &\geq 0, & (15)\\ R - pD_m &\geq 0, & (16)\\ Y - D_m &\geq 0, & (17)\\ D_m &\geq 0. & (18) \end{aligned}$$

These are the optimal information constraint (14) (in which D_m has a positive sign), the participation constraint for investors with a good signal (who attach probability q to success) (15), a non-participation constraint for uninformed investors (16), plus limited liability of the intermediary (17) as well as of investors (18).[25]

The intermediary's problem has the following solution:

Proposition 2 ***(The monitoring contract).*** *The intermediary offers a uniform contract with pay-off D_m in case of success, leading to a fraction of k_m informed investors, where*

$$\begin{aligned} D_m &= \frac{1}{2q} [qY + R], & (19)\\ k_m &= \frac{1}{2\overline{S}} u [qY - R]. & (20) \end{aligned}$$

Proof. *The solution for D_m follows from the only F.O.C.: $qY + R - 2qD_m = 0$. None of the constraints (15) to (18) bind. Constraint (18) cannot bind without violating (15). Neither of (17) or (15) can bind, as both would imply zero profit. Substituting the solution for D_m into (14) yields the solution for k_m.* ■

[25] Constraint (16) can be ignored. It only binds at values of $Y \geq (2q-1)R/pq$ at which the monitoring contract is not offered, as we will show below.

Under the monitoring contract, as (19) shows, the expected return to depositors, qD_m, is the average of expected returns on the risky asset (qY) and the safe asset (R). The monitoring contract thus implements some profit sharing between the intermediary and depositors. This makes it a combination of debt and equity similar to financial contracts often found in venture capital finance.

4.4.2 The optimal banking contract

Under a banking contract the intermediary gets money from all investors after a good signal (probability u), and from fraction $(1 - k_b(D_b))$ investors after a bad signal (probability $(1 - u)$). Fraction k_b is a *de*creasing function of (D_b), as the banking contract is an exit-game. Depending on the signal, success has probability q or $(1 - q)$. The intermediary thus solves

$$\max_{D_b} \Pi_b = uq[Y - D_b] + (1 - k_b(D_b))(1 - u)(1 - q)[Y - D_b], \tag{21}$$

subject to the following constraints:

$$\begin{aligned} k_b = \frac{V_b}{\overline{S}} &= \frac{1}{\overline{S}}(1 - u)[R - (1 - q)D_b], &(22)\\ pD_b - R &\geq 0, &(23)\\ qD_b - R &\geq 0, &(24)\\ R - (1 - q)D_b &\geq 0, &(25)\\ Y - D_b &\geq 0, &(26)\\ D_b &\geq 0. &(27) \end{aligned}$$

Note that in the information constraint (22) D_b has a negative sign. There are three participation constraints, one for each of the two groups of investors who accept the contract plus a non-participation constraint for investors with a bad signal.[26] Wealth constraints are identical to those under the monitoring contract.

The intermediary's problem has the following solution:

Proposition 3 ***(The banking contract).*** *The optimal banking contract has pay-off D_b in case of success, leading to a fraction of k_b informed in-*

[26] Constraint (25) can be neglected. It would only bind in the trivial case where $Y \gg R/(1 - q)$.

vestors, where

$$D_b = \frac{1}{p}R, \tag{28}$$

$$k_b = \frac{1}{\overline{S}}u\frac{q-p}{p}R. \tag{29}$$

Proof. *Participation constraint (23) binds: The intermediary must pay at least $D_b = R/p$ to attract uninformed investors. Higher D_b would mean lower profits, as the direct negative effect on profit from higher D_b is stronger than the indirect effect via lower k_b, i.e. $\partial\Pi_b/\partial D_b\big|_{D_b>R/p} < 0$, as follows from (22), (21), and from $(1-u)(q+p-1) = u(q-p)$. The latter equation is also used to solve for k_b after substituting R/p for D_b in (22).* ■

Under the banking contract the intermediary maximizes profits by setting the margin per depositor as high as possible (pay $D_b = R/p$) at the cost of having a smaller fraction expected depositors (more investors who get the potentially bad signal) than at higher levels of D_b. She sets uninformed investors to their participation constraint (23). As a consequence, the promised return on the banking contract D_b, unlike D_m, is independent of the potential return on the bank's asset, Y, a feature that reminds of debt.[27]

[27] We finally prove that the intermediary does not hold the riskfree asset. We show that the fraction of her funds invested in the riskfree asset, ρ ($0 \leq \rho \leq 1$), has an optimum value $\rho^* = 0$.

By holding the safe asset the intermediary could pay depositors ρR in case of failure, permitting her to pay only $D_b(\rho) = [1-(1-p)\rho]R/p$ in case of success. This would reduce the fraction of informed investors to

$$k_b(\rho) = (1-\rho)\frac{1}{\overline{S}}u\frac{(q-p)}{p}R = (1-\rho)k_b(0).$$

However, the direct loss from investing in the safe, rather than in the risky asset, exceeds the indirect gains from lower k_b. Rewriting (21) by use of optimal payments to depositors ($D_b(\rho)$ and ρR) as well as by weighting profits with $(1-\rho)$, taking derivatives, and rearranging by use of $\partial k_b(\rho)/\partial\rho = -k_b(\rho)/(1-\rho)$ yields the F.O.C.

$$k_b(\rho^*) = \frac{p}{2(1-q)(1-u)}.$$

The right hand side is bigger than unity, as follows from $p \geq 0.5 > 1-q$ and from (5). Further, (10), $k_b(\rho^*) > 1$, and $\overline{S} \geq u(q-p)R/p$ imply a negative solution for ρ^*. But negative ρ is not economically feasible (the intermediary cannot borrow at the riskfree rate), hence $\rho^* = 0$ ■.

4.5 Contract regions

We have derived the optimal monitoring contract (19) and the optimal banking contract (28). Among both, the intermediary chooses the one that maximizes her expected profit. We thus compare profits from the two rival contracts. Profits under the two optimal contracts (found by substituting optimal values for D_m, k_m and D_b, k_b, respectively, into (13) and (21)) are:

$$\Pi_m = \frac{1}{4\overline{S}} u^2 \left[qY - R\right]^2, \tag{30}$$

$$\Pi_b = \left[pY - R\right] \left[1 - \frac{1}{\overline{S}} u \left(1-u\right) \left(1-q\right) \frac{q-p}{p} R\right]. \tag{31}$$

The critical value of Y, at which the banking and the monitoring contract break even, called $\hat{Y}$, is the solution to a quadratic equation in $\hat{Y}$ which has one economically relevant solution.[28] It can be shown that for any admissible values of parameters R, $\overline{S}$, p, and q, the critical value $\hat{Y}$ lies in a particular interval corresponding to the left part of region B in Figure 2:

Lemma 4 *The border-line between the monitoring and the banking contract, $\hat{Y}$, satisfies*

$$\hat{Y} \in \left(\frac{1}{p} R, \frac{2q-p}{pq} R\right). \tag{32}$$

Proof. *The lower bound: If $Y = R/p$, $\Pi_b = 0$ and $\Pi_m > 0$.*
The upper bound: Assume that at $Y' = (2q-p)R/qp$, $\Pi_m > \Pi_b$. Setting $Y = Y'$ in (30) and in (31), using $\Pi_m > \Pi_b$ and rearranging yields $\overline{S} < u(q-p)R/p$, which contradicts assumption (10). ■

The two boundary values for $\hat{Y}$ are the closest values for which profits from both contracts can unambiguously be compared without knowledge of

[28]Setting

$$a = u^2 q^2 / 4\overline{S}, \text{ and } d = u\left(1-u\right)\left(q-p\right)\left(1-q\right) R / p^2 \overline{S},$$

the solutions are

$$\hat{Y}_{1,2} = \frac{\left[p\left(1-d\right) + 2a/q\right] \pm \sqrt{p\left(1-d\right)\left[p\left(1-d\right) + 4a/q\right]}}{2a\left[\left(1-d\right) + a/q^2\right] R}.$$

As all coefficients, including $(1-d)$, are positive, it is easy to show (by comparing squared expressions) that the numerator must be positive even when the negative sign of the square root applies. Therefore $\hat{Y}_{1,2} > 0$. The larger solution can be shown to violate (16); it is thus not economically relevant.

the actual values of R, $\overline{S}$, p, and q. At the lower bound, expected profit from the monitoring contract is positive, while profit from the banking contract is zero. At the upper bound, deposit returns and monitoring levels under both contracts are identical ($D_m = D_b$ and $k_m = k_b$), but the banking contract always attracts more depositors (the uninformed) and thus generates higher profits. Between the two boundary values, the monitoring or the banking contract can be more profitable, depending on model parameter values.

Inspection of (30) and (31) reveals the influence of different parameters on the choice of contracts. First, higher information costs favor the banking contract: With increasing $\overline{S}$, Π_m diminishes while Π_b increases. Second, an increase in p ceteris paribus favors the banking contract. Both effects tell the same story: A less risky asset (higher p) or a less reliable signal (lower q) work in the same direction as higher information costs; they make information of a certain quality less useful or more expensive and thus increase the profitability of the banking contract. Banks are often said to finance "opaque" assets. Opaqueness, in the language of this model, is either high $\overline{S}$ (information expensive) or small excess of q over p (information unreliable). In fact, opaque assets in this interpretation tend to be financed through banking contracts, rather than through monitoring contracts.

5 Comparison of monitoring levels

Having derived optimal contracts we can now compare monitoring levels with intermediation and without. In Figure 3 solid lines represent monitoring levels under the monitoring contract, k_m, and under the banking contract, k_b. Dotted lines, k_A and k_B, (borrowed from Figure 2) stand for monitoring levels in the non-intermediated equilibrium, when Y is in regions A ($Y < R/p$) and B ($Y \geq R/p$), respectively.

The monitoring contract is always offered when $Y < R/p$ and sometimes when $Y \geq R/p$. In the first case it unambiguously leads to a monitoring deficit ($k_m < k_A$). In the second case it may lead to deficient monitoring ($k_m < k_B$) or to excess monitoring ($k_m > k_B$), depending on whether $\hat{Y}$ lies to the left or to the right of the intersection of k_m and k_B. In contrast, the banking contract is never offered when $Y < R/p$ (in region A), as follows from Lemma (4). This allows us to state our main result in the form the following proposition:

	I_{Yg}	I_{Yb}	k
non-intermediated equ.			
– region A	k_A	0	$k_A = \frac{1}{S} u \left[qY - R \right]$
– region B	1	$1 - k_B$	$k_B = \frac{1}{S} (1-u) \left[R - (1-q) Y \right]$
intermediated equ.			
– monitoring contract	k_m	0	$k_m = \frac{1}{2S} u \left[q\overline{Y} - R \right]$
– banking contract	1	$1 - k_b$	$k_b = \frac{1}{S} (1-u) \frac{q+p-1}{p} R$

Table 3: Investment into risky asset and monitoring levels in different equilibria

Proposition 5 ***(Excess monitoring under the banking contract.)*** *The banking contract leads to excess monitoring, i.e. to*

$$k_b > k_B.$$

Proof. *From (32) $Y < R/p$; from (12), (8), and $D \leq \overline{Y}$, $k_b > k_B$.* ■

Banks, in the sense of firms borrowing from unsophisticated and thus uninformed investors, are excessively monitored (by a suboptimally high fraction of informed investors). As noted above, the monitoring level under the banking contract does not depend on Y, but on D. In contrast, monitoring in the non-intermediated equilibrium in region B falls in Y. Excess monitoring, as measured by the difference $k_b - k_B$, thus increases in Y and reaches its maximum at $R/(1-q)$, where investors with direct access to the risky asset would invest blindly.

[Figure 3 about here]

Resources spent on excess monitoring are not fully lost. More monitoring does not only mean more (deadweight) signal expenditure, but also a higher fraction of informed investors, i.e. less investment in the risky asset after a bad signal (which has negative expected present value). Using aggregate expected return as a social welfare criterion we show that the excess signal expenditure nevertheless exceeds the resources saved by less investment after a bad signal.

Proposition 6 ***(Welfare loss from excess monitoring.)*** *The level of welfare is lower under the banking contract than under the corresponding*

non-intermediated equilibrium.
Proof. *Welfare levels (net of signal expenditure) in the non-intermediated equilibrium (Region B) and under the banking contract, W_B and W_b, are:*

$$\begin{aligned} W_B &= pY + \frac{1}{2\overline{S}} (1-u)^2 [R - (1-q) Y]^2 , \\ W_b &= pY + \frac{1}{2\overline{S}} (1-u)^2 \left\{ [R - (1-q) Y]^2 - \left(\frac{1-q}{p} \right) [pY - R] \right\} . \end{aligned}$$

As the banking contract is only offered when $pY > R$, it follows that $W_B > W_b$. ■

This means that the expected return saved by smaller investment in the risky asset after a bad signal does not fully compensate excessive monitoring expenditure. Excess monitoring thus implies a welfare loss. More generally, welfare is lower in the intermediated equilibrium than in the non-intermediated equilibrium.[29]

6 Concluding Remarks

We have shown in the context of financial intermediation that in models of endogenous information it might be crucial to distinguish between assets that are attractive to uninformed investors (like bank deposits), and assets that are not (credits to non-banks). We found that a reduction in promised returns diminishes investors' demand for information in the former case, and increases it in the latter. This finding may have more general applications. It may explain information demand in the job market or in international lending, to take two examples. Yet, we think it is particularly important

[29] This is also true under the monitoring contract. We limit ourselves to give the lines of proof for the two possible cases:

(1.) In region A ($R/q < Y < R/p$), the monitoring contract implies insufficient monitoring ($k_m = k_A/2$). With the same logic used in the text for the banking contract it can be shown that $W_m < W_A$, as the signal expenditure saved by insufficient monitoring falls short of the expected return lost through smaller investment in the project after a good signal.

(2.) In region B ($Y \geq R/p$) the monitoring contract may lead to insufficient or to excess monitoring. Yet, as $W_m < W_A$ and, in this region, $W_A < W_B$, it follows that $W_m < W_B$. The monitoring contract implies lower welfare than the non-intermediated equilibrium in this region as well.

in the area of financial intermediation. It means that banks are excessively monitored. Investors spend more money on information when they consider lending to a bank than when they could directly hold the bank's assets.

Our paper tells a story about the delegation cost of intermediation. We found a distortion of investors' monitoring decisions in the presence of asymmetric information and limited liability. In the absence of perfect diversification by the intermediary (as assumed by Diamond, 1984) and of perfect competition between intermediaries, delegation is costly. Excess monitoring is one form in which delegation costs occur.

The main prediction from our model, excess monitoring of banks, contrasts with the predominant "deficient monitoring view" formulated by Dewatripont and Tirole (1994).[30] Both models provide complementary partial insights into endogenous monitoring of an intermediary by investors. Dewatripont and Tirole (1994) focus on the public good character of information. Our model examines the impact of intermediation on the demand for information. Both models thus describe potentially important aspects of financial intermediation.

Their empirical and policy implications differ from ours. Dewatripont and Tirole expect banks to hold "low monitoring assets". In our model, the banking contract tends to be optimal for the intermediary if it holds a risky asset with high a priori probability of success (high p), a weak signal (low q), or an expensive signal (high $\overline{S}$). Such assets are relatively safe, but informationally "opaque". Evidence that banks are indeed opaque is found by Morgan (1999). While Dewatripont and Tirole do not explicitly model intermediaries' liability structure, our model also explains features of contracts found in practice. The promised return on the banking contract is independent of the potential return on the bank's asset, reminding of debt. In contrast, the monitoring contract is a combination of debt and equity, reminding of venture capital finance.

Contrary to Dewatripont and Tirole, we have assumed that information cannot be communicated from one agent to the other. This is a strong assumption. In reality, investors may have both kinds of information: (1.) knowledge that may be difficult to share due to communication costs or to credibility problems, and (2.) information that travels easily, or that benefits the uninformed through actions of the informed (e.g. through liquidating

[30] In reality, banks may also be insufficiently monitored due to (excessive) implicit or explicit deposit insurance.

bad banks, as in Calomiris and Kahn, 1991, or in Park, 2000). Both, the Dewatripont and Tirole model and our own, thus explain some aspects of intermediation. Only a model allowing for both kinds of information could probably replicate the rich variety of existing financial institutions, including banks, non-bank intermediaries, as well as rating agencies and bank supervisors.

Appendix

We prove that the intermediary does not hold the riskfree asset. We show that the fraction of her funds invested in the riskfree asset, ρ ($0 \leq \rho \leq 1$), has an optimum value $\rho^* = 0$.

By holding the safe asset the intermediary could pay depositors ρR in case of failure, permitting her to pay only $D_b(\rho) = [1 - (1-p)\rho] R/p$ in case of success. This would reduce the fraction of informed investors to

$$k_b(\rho) = (1-\rho) \frac{1}{S} u \frac{(q-p)}{p} R = (1-\rho) k_b(0).$$

However, the direct loss from investing in the safe, rather than in the risky asset, exceeds the indirect gains from lower k_b. Rewriting (21) by use of optimal payments to depositors ($D_b(\rho)$ and ρR) as well as by weighting profits with $(1-\rho)$, taking derivatives, and rearranging by use of $\partial k_b(\rho) / \partial \rho = -k_b(\rho) / (1-\rho)$ yields the F.O.C.

$$k_b(\rho^*) = \frac{p}{2(1-q)(1-u)}.$$

The right hand side is bigger than unity, as follows from $p \geq 0.5 > 1-q$ and from (5). Further, (10), $k_b(\rho^*) > 1$, and $S \geq u(q-p) R/p$ imply a negative solution for ρ^*. But negative ρ is not economically feasible (the intermediary cannot borrow at the riskfree rate), hence $\rho^* = 0$ ■.

References

Allen, Beth, 2000, The Future of Microeconomic Theory, *Journal of Economic Perspectives*, 14, 143-150

Birchler, Urs W., 2000, Bankruptcy Priority for Bank Deposits: A Contract Theoretic Explanation, *Review of Financial Studies*, 13, 813-839.

Cukierman, Alex, 1980, "The Effects of Uncertainty on Investment under Risk Neutrality with Endogenous Information", *Journal of Political Economy*, 88, 462-475.

Dewatripont, Matthias. and Jean Tirole, 1993, *La Réglementation prudentielle des banques*, (Conférence Walras-Pareto lectures), HEC Univ. de Lausanne, Payot, Lausanne.

Diamond, Douglas, 1984, "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, 51, 393-414.

Diamond, Douglas W. and Phillip Dybvig, "Bank Runs, Deposit Insurance, and Liquidity", *Journal of Political Economy*, 91, 401-419.

Dow, James and Nathalie Rossiensky, 1998, "The Cost of Debt for a Financial Firm", working paper, London Business School.

Freixas Xavier, and Jean-Charles Rochet,1997, *Microeconomics of Banking*, MIT Press, Cambridge, Mass.

Garcia, Gillian, 1999, "Deposit Insurance: A Survey of Actual and Best Practices", IMF working paper 99/54.

Greenbaum, Stuart I. and Anjan V. Thakor, 1995, *Contemporary Financial Intermediation*, The Dryden Press, Orlando, Fla.

Hellwig, Martin, 1998, Allowing for Risk Choices in Diamond's "Financial Intermediation as Delegated Monitoring", working paper, University of Mannheim.

Jackson, Thomas. H. and A. T. Kronman, 1979, "Secured Financing and Priorities Among Creditors", *Yale Law Journal*, 88, 1105-1182.

Krasa, Stefan, and Anne Villamil, 1992, Monitoring the Monitor: An Incentive Structure for a Financial Intermediary," *Journal of Economic Theory,* 57, 197-221.

Morgan, Donald P., 1999, "Judging the Risks of Banks: Why Can't Bond Raters Agree?", working paper, Federal Reserve Bank of New York.

Park, Cheol, 2000, "Monitoring and the Structure of Debt Contracts", *The Journal of Finance*, 55, 2157-2193.

Wallace, Neil, 1988, "Another Attempt to Explain an Illiquid Banking System: The Diamond-Dybvig Model with Sequential Service Taken Seriously", *Quarterly Review of the Federal Reserve Bank of Minneapolis*, 12, 3-16.

Xu, Bin, 2000, The Welfare Implications of Costly Monitoring in the Credit Market: A Note, *The Economic Journal*, 110, 526-580.

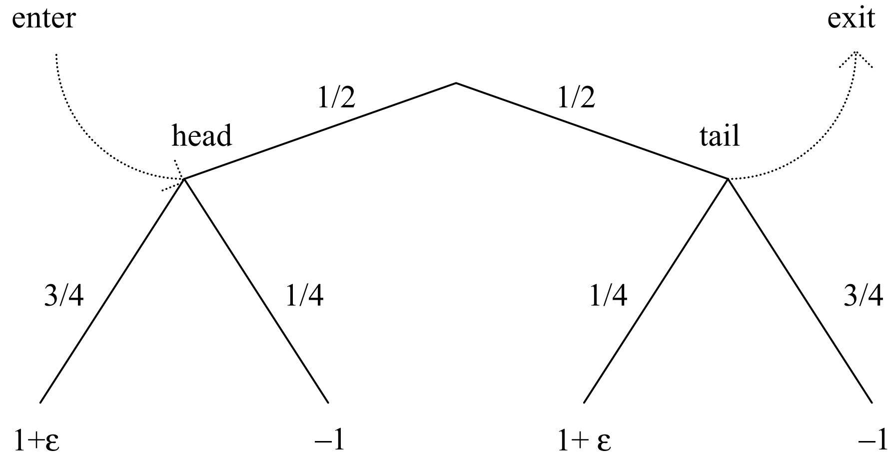

Figure 1: Neil's Problem

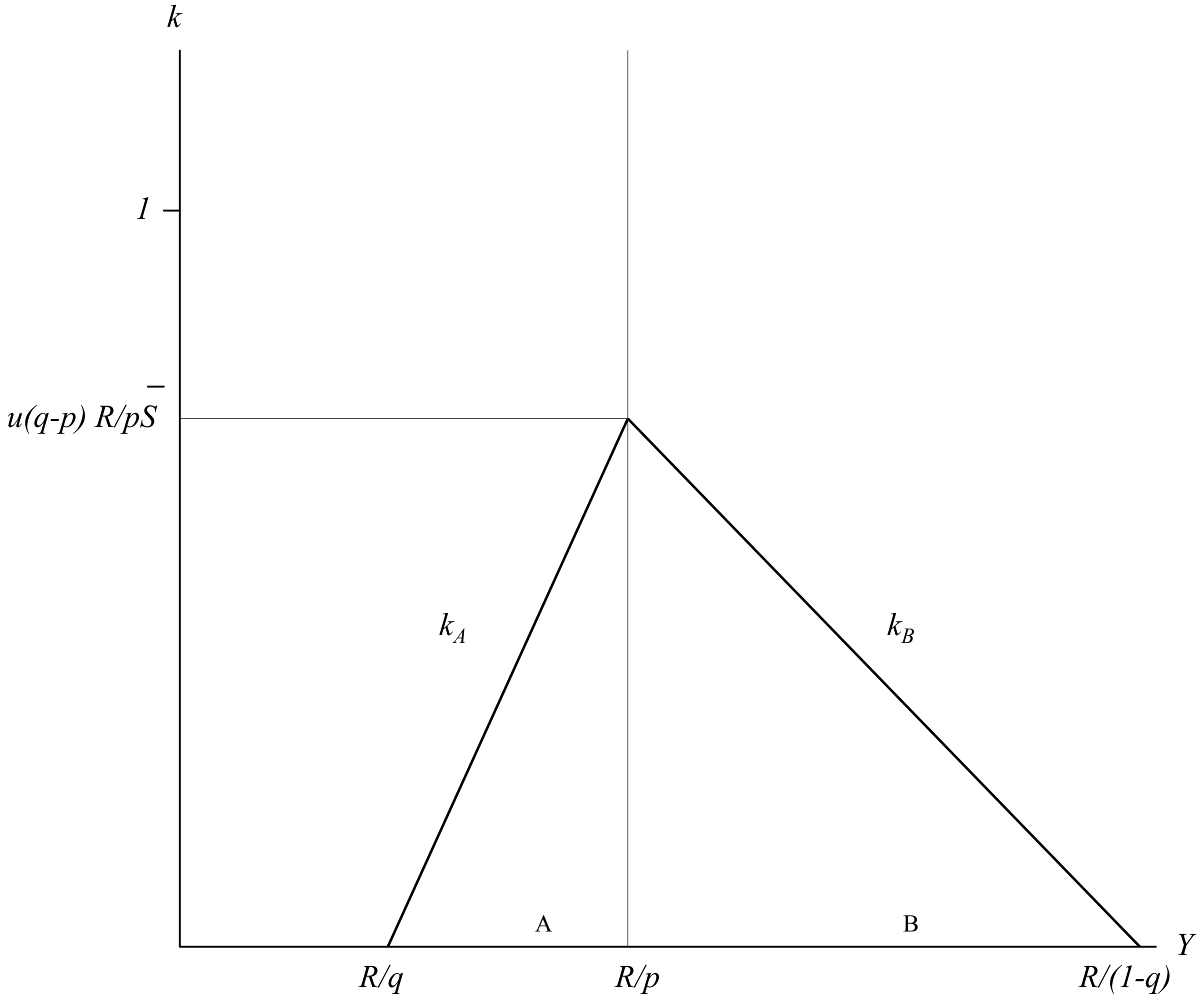

Figure 2: Monitoring in the non-intermediated equilibrium

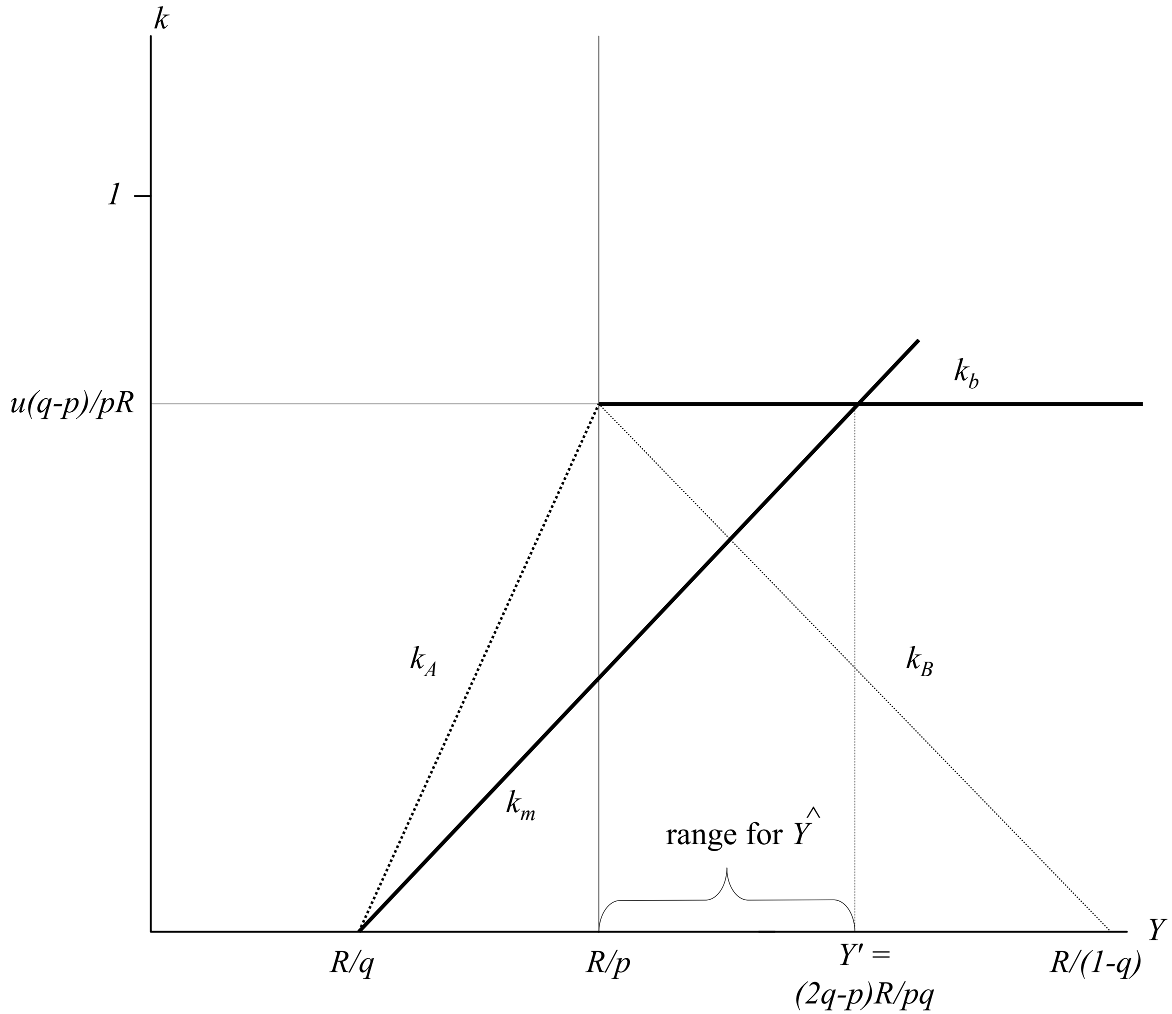

Figure 3: Monitoring in the intermediated equilibrium