Financial intermediation and the optimal tax system

Ramon Caminal
Institut d’Anàlisi Econòmica, CSIC, 08193 Bellaterra, Barcelona, Spain
Received May 1994 final version received November 1995

Abstract

In this paper a stylized general equilibrium model is constructed to analyze the relative efficiency of taxing financial intermediaries. A crucial feature of the model is that tax collection costs are endogenous, i.e. they result from primitive assumptions about information and transaction costs, instead of being assumed ad hoc. The model provides useful insights into the welfare costs and incidence of banks’ reserve requirements or, equivalently, of a tax on deposits. In particular, it is shown that a tax on bank deposits can be part of an optimal tax system, provided banks’ monopoly power is significant. However, if the banking industry is perfectly competitive, the first dollar of revenue is more efficiently collected by a uniform capital income tax, although a strictly positive welfare loss is incurred.

Keywords: Reserve requirements; Financial intermediation; Optimal taxes; Tax collection costs

JEL classification: G21; H21

1. Introduction

Is it efficient to finance government spending through a minimum reserve requirement on banks? More generally, can a tax on the activity of financial intermediaries be part of an optimal tax system, or is it likely to be dominated by alternative taxes? To analyze these issues rigorously, we develop a stylized general equilibrium model of banking, in which the information-gathering costs of market participants and tax authorities are treated explicitly and symmetrically.
In many countries the most important tax levied on financial intermediaries are reserve requirements. The fact that reserve requirements are an implicit form of taxation has long been recognized (see, for instance, Fama, 1980). Such a requirement induces banks to hold a higher fraction of their deposits in the form of non-interest-bearing reserves, which reduces the average return of the banks' portfolios and increases their demand for monetary base. Thus, the reserve requirement operates as a tax by increasing seigniorage revenue (by expanding the base of the inflation tax). More specifically, under certain assumptions a reserve requirement is equivalent to a proportional tax on deposits plus an open market sale of government bonds of an amount equivalent to the volume of resources kept captive by the requirement (Romer, 1985, and Bacchetta and Caminal, 1994).

While it could be argued that the government sets reserve requirements mainly to facilitate monetary control or prevent bank runs, the following three observations suggest that very often revenue-raising is the most important objective. First, in many countries required reserve ratios are very high. It is difficult to argue that a reserve ratio of an order of magnitude of 20% or 30% is needed to stabilize any monetary aggregate or to compensate for the suboptimally low level of voluntary reserves. Secondly, required reserve ratios are positively correlated with inflation rates, which is compatible with the minimization of the welfare costs of inflationary finance (Brock, 1989).\(^1\) Thirdly, a negative correlation seems to exist between reserve ratios and per capita GDP, which suggests that countries with less developed tax collection systems rely more heavily on such forms of taxation.\(^2\)

Informal discussions about the potential optimality of reserve requirements have focused, on the one hand, on the distorting effects of such a tax (the disintermediation effect) and, on the other hand, on the low tax collection costs. It is well known that, in competitive environments, taxing intermediate goods is usually dominated by taxing final goods (Diamond and Mirrlees, 1971). In the case of financial intermediaries, the intuition behind this result is clear. Presumably, these intermediaries exist because they perform an efficiency-enhancing role in the allocation of savings. Thus, taxing their activity distorts an additional margin. It induces some savers to avoid intermediaries and to invest their funds directly (the disintermediation

\(^1\) However, the degree of substitutability between reserve requirements and inflation rates varies with the degree of international capital mobility (Bacchetta and Caminal, 1992).

\(^2\) Additionally, it is often argued that a reserve requirement is likely to be a redundant instrument of stabilization policy. See, for instance, Horrigan (1988).
effect). Therefore, in principle, taxing financial intermediaries should be dominated by taxing final investment projects.  

Alternatively, it is often argued that tax collection costs may be substantially higher for individual production units than for large financial intermediaries (the tax collection costs effect). In other words, it may be much easier for the government to monitor a few large financial institutions than a large number of individual entrepreneurs. This is how reserve requirements are usually justified in countries with less developed financial and taxation systems.

In this paper we consider a model of the credit market that focuses on such a trade-off. We assume that all agents are ex ante identical, and hence the natural social welfare criterion is the maximization of ex ante utility. As idiosyncratic uncertainty resolves, agents take labor allocation decisions: either they become entrepreneurs (and borrow in the credit market), or workers (and net lenders). Production requires two inputs in fixed proportions: the entrepreneur's labor and external finance (capital). Entrepreneurs can choose to finance their investment projects either directly from savers (issuing bonds) or through financial intermediaries (applying for a bank loan). Thus, taxing financial intermediaries may cause a disintermediation effect, as some loan applicants are diverted to the bond market. By monitoring investment decisions financial intermediaries reduce agency costs; and hence, disintermediation is welfare decreasing. Finally, the government faces the same informational asymmetries as does the private sector and, consequently, the collection costs associated with certain taxes arise endogenously. In other words, the government is assumed to have neither an informational advantage nor a disadvantage vis-à-vis the private sector.

We show that, with a competitive banking industry, the most efficient way of collecting a small amount of revenue is through a uniform capital income

---

3 A particular application of this principle is contained in Kimbrough (1989), which is the only attempt I know of to formally analyze the potential optimality of reserve requirements from the taxation point of view. He assumes that bank deposits (as well as cash) reduce shopping time, and hence they are an argument of the production function. In his framework, the reserve requirement falls exclusively on depositors, and therefore it is a tax on an intermediate good and dominated by a tax on final consumption.

4 Exogenous tax collection costs have been extensively considered in the literature (see, for instance, Slemrod, 1990, and references cited there). That is to say, independently of the behavior of the private sector, it is assumed that different taxes carry different collection costs. This ad hoc approach is reasonable if collection costs are purely administrative. However, if they are related to the informational requirements for implementing the tax scheme, then important insights can be gained by placing both the government and the private sector on a level playing field.
A tax on output (i.e., on the return of final investment projects) creates no allocative distortion, because it falls proportionally on both inputs (capital and entrepreneurial activity), but involves a fixed tax collection cost. Alternatively, a uniform capital income tax distorts the allocation of resources and raises the interest rate, which increases the frequency of bankruptcies. The first effect vanishes for the first dollar collected, but the second is strictly negative, even for negligible tax rates. Finally, a tax on bank deposits has similar negative effects to the uniform capital income tax and, additionally, it induces disintermediation.

Next, we ask whether the competitive structure of the banking industry plays any role in the determination of the optimal tax system. The answer is affirmative. In the extreme case of a completely segmented banking system, banks charge entrepreneurs their reservation price (the rate they would have to pay in the bond market). Thus, a tax on bank deposits cannot be passed on to the loan rates, and is paid exclusively by the bank. The only effect of this tax is now the disintermediation effect, but the welfare cost of this effect is zero for the first dollar collected, and therefore in this case the most efficient tax scheme involves a positive tax on bank deposits.

The paper is organized as follows. Section 2 describes the model. The properties of the competitive equilibrium are analyzed in Section 3. Section 4 deals with the effect of linear taxes in the case of a competitive banking industry. Section 5 considers the effect of monopoly power in banking. Finally, some concluding remarks close the paper.

2. A simple general equilibrium model with financial intermediation

In this section we construct the simplest possible model that embeds the following features: (a) interpersonal welfare comparisons are not needed; (b) credit market conditions affect the allocation of real resources; (c) the supply of and demand for credit are not perfectly inelastic; and (d) firms may choose between bank loans and an alternative financial instrument. Some of the assumptions we make may appear to be arbitrary, but actually allow the model to fulfill these requirements and still remain tractable.

\[5\] To collect the output tax the government has to monitor entrepreneurs under all circumstances.

\[6\] The reason is that the equilibrium is ex ante inefficient, even with perfectly competitive markets and no government intervention. The intuition is that the interest rate that clears the market is too high, in the sense that it induces too many resources spent on bankruptcy procedures.

\[7\] For simplicity, we consider a direct tax on deposits instead of a reserve requirement. It is unlikely that we would learn much here by worrying about the composition of the government's liabilities.
The model is static, in the sense that no real time elapses but decisions take place sequentially. There are three goods: labor, an intermediate good (capital), and a consumption good. There is a continuum of ex ante identical agents, indexed by \( i \), who derive utility only from the consumption good and are risk-neutral (they maximize expected consumption). Agents are endowed with one unit of labor that they can allocate to alternative activities. On the one hand, agent \( i \) has potential access to an investment project that produces \( y_i \) units of the consumption good from one unit of the intermediate good and one unit of own labor (agent \( i \) can become an entrepreneur who eventually will need external finance). The development of such an investment project is done in two stages, at each stage the entrepreneur employs half a unit of labor. On the other hand, all agents have free access to a constant returns to scale technology that produces one unit of the intermediate good per unit of labor (the agent can become a 'worker', and eventually a supplier of funds).

The output of individual \( i \)'s project, \( y_i \), is the sum of two random variables, \( w_i \) and \( x_i \):

\[
y_i = w_i + x_i.
\]

(1)

The timing is as follows (see Fig. 1):

- **Stage I: First occupational choice.** After a signal of \( w_i \), denoted by \( s_i \), is publicly observed, agents choose either to employ their whole labor endowment in the intermediate good sector (become a worker), or to employ half a unit of labor in developing the first stage of the investment project (become an entrepreneur).\(^8\)

- **Stage II: Financial market.** Once the random variable \( w_i \) is realized and publicly observed, workers and entrepreneurs meet and sign financial contracts, establishing the terms at which the intermediate good is exchanged for the consumption good.

- **Stage III: Second occupational choice.** According to the contracts they signed, entrepreneurs either abandon their project and use the remaining half unit of labor in producing the intermediate good, or they invest their remaining half unit of labor in developing the second stage of the project. Those entrepreneurs in their second stage choose an action \( a_i \) that

---

\(^8\) Thus, the decision to become a worker is irreversible, but engaging in entrepreneurial activity is partially reversible, as only half unit of labor is committed at this point.
influences the distribution of \( x_i \). Such an action can be monitored by outsiders at a cost.

- **Stage IV: Contract implementation.** The intermediate good is allocated among investment projects. Production takes place and the random variable \( x_i \) is realized. Finally, output, which might be monitored by outsiders at a cost, is distributed.

The timing of the financial market stage turns out to be crucial. We rule out the possibility of signing contracts before any uncertainty is resolved (at the beginning of Stage I), i.e. we assume that markets are incomplete. Also, we let agents sign contracts before all labor decisions have been taken in order to allow for elastic supply and demand functions and have a unique equilibrium interest rate.\(^9\) In other words, if entrepreneurs sign contracts after committing their second half unit of labor, then the supply of and demand for intermediate good are inelastic and, consequently, the equilibrium price is not well defined.

In our simple world the only potential exchanges are of the intermediate good for claims to final output (we call this credit). On the supply side are the workers, and on the demand side the entrepreneurs. The occupational choice implies that agents select which side of the credit market to be on. On the one hand, this reflects a real-world situation: when the size of available investment projects exceeds personal funds, individuals must choose whether to borrow and directly manage the investment project or to lend to other potential entrepreneurs. On the other hand, modeling individuals as being ex ante identical will allow us to illustrate the distortions caused by capital income taxation in a simple framework, which conveys a natural social welfare criterion (the maximization of aggregate consumption).

The random variables \( w_i \) and \( s_i \) are jointly distributed according to the density function \( f(w, s) \) on \([0,1] \times [0,1]\), and \( x_i \) is distributed according to \( h(x|a) \) on \([0,1]\). Action \( a_i \) can take two values: \( a_0 \) and \( a_1 \). If we let capital letters denote distribution functions, then the assumptions on \( f(\cdot, \cdot) \) and \( h(\cdot) \) can be stated as follows:

**Assumption 1.** Positive density in support, i.e. \( f(w, s) > 0 \) for all values of \( w \) and \( s \) in \([0,1]\), and \( h(x|a) > 0 \) for all \( x \) in \([0,1]\) and \( a_j, j = 0, 1 \). Both \( f(w, s) \) and \( h(x|a) \) are continuously differentiable in all their arguments.

**Assumption 2.** The signals \( s \) order the distributions of \( w \) according to first-order stochastic dominance; in particular \( (dF(w|s)/ds) < 0 \) for all \( s \).

**Assumption 3.** Action \( a_0 \) implies the same expected return but higher risk.

\(^9\) Alternatively, we could have allowed for variable size investment projects and let contracts be signed after all labor decisions are taken. Unfortunately, such a version of the model is much less tractable.
than action $a_1$. More specifically, $E(x|a_0) = E(x|a_1) = 1/2$ and $h(x|a)$ are symmetric around $1/2$. There is an $x_0$, $0 < x_0 < 1/2$, such that

\begin{align*}
    h(x|a_0) > h(x|a_1), & \quad \text{if } 0 \leq x < x_0, \\
    h(x|a_0) < h(x|a_1), & \quad \text{if } x_0 < x < 1 - x_0, \\
    h(x|a_0) > h(x|a_1), & \quad \text{if } 1 - x_0 < x \leq 1.
\end{align*}

Since all random variables are independent and there is a continuum of agents, there is no aggregate uncertainty. Assumption 2 implies that the higher the value of $s$ the higher is the conditional expected value of $w$.$^{10}$ Assumption 3 is stronger than second-order stochastic dominance (the distribution of output induced by action $a_1$ second-order stochastically dominates the distribution induced by $a_0$) and greatly simplifies the entrepreneurs' financial decision. Assumption 1 plays a minor technical role in the analysis.

The realizations of $s_i$ and $w_i$ are public information. Action $a_i$ is chosen by entrepreneur $i$ and is only observable by outsiders by using an ex ante monitoring technology. Such a technology requires spending $C$ units of the intermediate good in observing $a_i$. For simplicity, we assume that, after paying $C$, $a_i$ is verifiable, and hence contractable. Ex ante monitoring is carried out by financial intermediaries, and hence we will call financial contracts that involve ex ante monitoring 'bank loans'. In other words, in this model banks are special only because they are able to monitor firms' actions and thus increase the expected return of investment. Non-bank financial contracts will be called 'bonds' and involve no ex ante monitoring.

Similarly, the entrepreneur costlessly observes the realization of $x_i$, but outsiders have to pay $D$ units of the consumption good to observe $x_i$ and produce hard evidence. Such an ex post monitoring technology can be interpreted as a legal bankruptcy procedure. In those circumstances made explicit in financial contracts, lenders file the firm for bankruptcy, authorities automatically monitor the firm's assets, liquidates them and distributes the proceeds among the claimants. Under this interpretation the number of financiers per firm is irrelevant, since total monitoring costs are independent of the number of claimants.$^{11}$ Also, we assume away the

---

$^{10}$ Integrating by parts, we can write the conditional expected value of $w$ as

$$E(w|s) = 1 - \int_0^1 F(w|s) \, dw.$$ 

Thus,

$$\frac{dE(w|s)}{ds} = -\int_0^1 \frac{dF(w|s)}{ds} \, dw > 0.$$ 

$^{11}$ Diamond (1984) argues that it may be efficient to delegate ex post monitoring to large financial intermediaries. We could accommodate this idea in our framework, but this would require modeling intermediation in the bond market explicitly.
possibility of renegotiation. As we will see below, the failure to meet their interest obligations automatically triggers the opening of such a bankruptcy procedure.

3. The competitive allocation

In this section we analyze the equilibrium allocation when markets are perfectly competitive. In our model of the credit market, workers (savers) supply the intermediate good (capital) in exchange for a share of the entrepreneurs' final output. Since output is a random variable and there is the possibility of ex ante and ex post monitoring, financial contracts can, in principle, be quite complicated. First, we analyze the structure of optimal contracts, with and without ex ante monitoring. Secondly, we turn to the entrepreneurs' financial choice. Thirdly, we describe how financial markets operate and characterize the equilibrium.

3.1. Optimal financial contracting without ex ante monitoring (bonds)

If parties ignore the possibility of using the ex ante monitoring technology, the optimal contract specifies the lender's payoff in each state of nature, and the circumstances under which ex post monitoring takes place. The literature on financial contracting with costly state verification (Townsend, 1979; Diamond, 1984; Gale and Hellwig, 1985) shows that, under those circumstances, the optimal contract is a standard debt contract. In our case this also implies that the entrepreneur is expected to choose the riskier action, \( a_0 \), since the entrepreneur's payoff function under a debt contract is convex. If we denote by \( S^b \) the lender's payoff function and by \( r^b \) the lender's required expected rate or return, the following proposition characterizes the structure of the optimal contract.

**Proposition 1.** Without ex ante monitoring, the optimal contract takes the following form (see Fig. 2):

\[
S^b = \begin{cases} 
R^b, & \text{if } w + x \geq R^b, \\
 w + x - D, & \text{otherwise}, 
\end{cases}
\]

and monitoring takes place if and only if \( w + x \leq R^b \). If \( w \geq r^b \), then \( R^b = r^b \), and if \( w < r^b \), then \( R^b \) is given by the lowest solution to

\[
R^b - w = \int_0^{R^b - w} (w + x - D)h(x|a_0) \, dx + R^b [1 - H(R^b - w|a_0)] = r^b.
\]
If \( r^b - w \) is sufficiently large, then Eq. (2) has no solution, i.e. for any \( R^b \leq w + 1 \) the lender’s expected rate of return is below \( r^b \). In other words, if \( w \) is sufficiently low, then a bond may not be able to generate the required rate of return.

The intuition behind Proposition 1 goes as follows. An incentive-compatible contract requires monitoring output in those states of nature in which the payment to the lender is strictly below the maximum. Also, optimality requires the minimization of monitoring costs. If the borrower’s payoff function is strictly convex, then the entrepreneur chooses the riskier action \( a_0 \). But if the borrower’s payoff function is not strictly convex, then the lender’s payoff function is strictly monotone, and hence output has to be monitored in all states. As a result, the optimal contract involves a fixed payment, \( R^b \), except when output is below \( R^b \), in which case the realization of \( x \) is monitored and the lender receives all the output (net of monitoring costs).\(^{12}\) Thus, in the optimal contract the entrepreneur’s payoff function is

\(^{12}\) The optimal contract can be thought of as being implemented by a mechanism that involves the borrower announcing the level of output, and the contract specifying the payment to the lender and the probability of monitoring as a function of the announcement. Note that if cheating has no exogenous cost, then incentive compatibility requires the probability of monitoring to be one for those states that involve payments strictly below the maximum level. Hence nothing is gained by allowing random monitoring. However, if there are exogenous cheating costs (established, for instance, by the legal system), then the efficiency of the contract improves by allowing random monitoring. In the simple case of these costs being a proportion \( \gamma \) of the difference between the true output level and the announcement, then the optimal contract involves a probability \( 1/(1 + \gamma) \) of monitoring in the default states, which reduces the absolute magnitude of monitoring costs without altering any of the qualitative results.
strictly convex and the riskier action \( a_0 \) is chosen. (See the appendix for a formal proof.)

Financing investment projects with \( r^b > w \) through bonds involves a deadweight loss. The reason is that the entrepreneur chooses the riskier action, but such a choice is inefficient (provided the probability of bankruptcy is lower than one-half) since bankruptcy costs are not minimized. To see this, we can write the borrower’s payoff as

\[
\pi^b(w) = \int_{R^b - w}^{1} (w + x - R^b)h(x|a_0) \, dx .
\]  

By substituting Eq. (2) into Eq. (3) we obtain:

\[
\pi^b(w) = w + \left(1/2\right) - r^b - DH(R^b - w|a_0). 
\]  

Thus, if the action were observable (and again if \( R^b - w < 1/2 \)), then the efficient choice would be \( a_1 \), since it is the action that minimizes expected bankruptcy costs. Note that the output loss associated with the asymmetric information problem (the agency cost) is a negative function of \( w \).

### 3.2. Optimal financial contracting with ex ante monitoring (bank loans)

Banks are the only agents that have access to the ex ante monitoring technology, i.e. a bank can observe the entrepreneur’s action \( a_i \) after paying \( C \) units of the intermediate good. In this case the contract could specify the payment to the bank and whether to monitor ex post or not as a function of both \( x_i \) and \( a_i \). Also, the contract can condition the funding of the project on the action observed.

Given the costly state verification problem, the optimal contract with ex ante monitoring is still a standard debt contract. The only difference between a bond and a bank loan is that the latter requires the entrepreneur to choose the efficient action, \( a_1 \). In other words, since ex ante monitoring is costly, a bank loan can only be optimal if it reduces expected bankruptcy costs by inducing the borrower to take the efficient action. If we denote by \( S^e \) the lender’s payoff in this type of contract, and the required rate of return is \( r^e \), then a bank loan is characterized by the following proposition.

**Proposition 2.** An optimal contract with ex ante monitoring consists of providing one unit of the intermediate good if and only if \( a = a_1 \), and

\[
S^e = \begin{cases} 
R^e , & \text{if } w + x \geq R^e , \\
w + x - D , & \text{otherwise} ,
\end{cases}
\]
and ex post monitoring takes place if and only if \( w + x \leqslant R^e \), and \( R^e \) is given by the lowest solution to:

\[
R^e - w \int_0^{R^e-w} (w + x - D)h(x|a_1)dx + R^e \left[ 1 - H(R^e - w|a_1) \right] = r^e(1 + C).
\]  

(5)

Analogously to Eq. (4), using Eq. (5) we can write the borrower's payoff as follows:

\[
\pi^e(w) = w + (1/2) - r^e - DH(R^e - w|a_1).
\]  

(6)

3.3. The entrepreneur's financial decision

If entrepreneurs are faced with the choice between bonds and bank loans, then they will consider the following trade-off. Bank loans may reduce expected bankruptcy costs by inducing the entrepreneur to take the efficient action; however, loan rates must cover the ex ante monitoring costs. Clearly, if agency costs are very high (if the output loss due to the inefficient action is high), then the entrepreneur will prefer to finance investment through a bank loan, but if agency costs are low, then a bond is preferred.

More specifically, an entrepreneur compares \( \pi^b(w) \) and \( \pi^e(w) \), which are given by Eqs. (4) and (6), respectively. Suppose that \( r^b = r^e = r \). If \( w \geq r \), then a bond is preferred to a loan because the probability of bankruptcy is zero and the choice of action is irrelevant. The choice is not so trivial when \( w \) is in the interval \([w_m, r]\), where \( w_m \) is the value of \( w \) that implies a probability of bankruptcy under a bond contract equal to 1/2. As shown by Lemma 1 in the appendix, in equilibrium there is no entrepreneur with \( w < w_m \). In this range the optimal financial decision is given by the following proposition.

Proposition 3. If \( C \) is small enough, then there exists a \( w_1, w_m < w_1 < r \), such that

\[
\pi^e(w) > \pi^b(w), \quad \text{if } w_m < w < w_1,
\]

and

\[
\pi^e(w) < \pi^b(w), \quad \text{if } w_1 < w < r.
\]
Also, \( \lim_{C \to 0} w_1 = r \).

If \( C \) is relatively large with respect to the economic value of the information obtained by monitoring the investment project, then a bank loan will never be efficient. Thus, to make the problem non-trivial, without making additional assumptions on the difference between \( h(x|a_0) \) and \( h(x|a_1) \), we require \( C \) to be sufficiently small.

If \( w \) is sufficiently close to \( r \), then since the probability of bankruptcy is very low, the output losses caused by choosing the riskier action are not compensated for by the ex ante monitoring costs, and a bond is preferred to a bank loan. Only when the probability of bankruptcy is significant is it efficient to monitor the action \( a_1 \), and thus an entrepreneur with a relatively low \( w \) will prefer a bank loan over a bond. As the ex ante monitoring cost, \( C \), goes to zero the probability of bankruptcy under a bond contract by the entrepreneur indifferent between the two financial instruments must also go to zero.\(^1\)

To simplify computations (without losing any economic insight) we assume that \( C \) is negligible. In other words, we only consider the limit of equilibria with \( C > 0 \), as \( C \) goes to zero. In this case, entrepreneurs with \( w \geq r \) borrow in the bond market, with \( R_b = r \), and get an expected payoff equal to \( w + (1/2) - r \), while entrepreneurs with \( w < r \) apply for loans. In the latter case the entrepreneur has an expected payoff equal to \( w + (1/2) - r - DH(R^e - w|a_1) \), with \( R^e \) given by Eq. (5).

3.4. Market equilibrium

After the random variables \( w \), are realized, agents meet to sign financial contracts. Workers and entrepreneurs can trade directly in a centralized bond market, or through banks. In the bond market only contracts without ex ante monitoring can be traded. In such a market there is an exogenous market-clearing mechanism.

There is also a large number of banks, who offer entrepreneurs contracts with ex ante monitoring (loans) and offer workers (depositors) a safe interest rate. They can also trade in the bond market. Banks set deposit and loan rates in order to maximize expected profits. They can offer a safe interest rate to their depositors by having a perfectly diversified loan portfolio.\(^2\)

The timing of the financial market stage is the following. First, banks

\(^1\) See Diamond (1991) for an alternative approach of the firms' choice between loans and bonds.

\(^2\) In this section we could have assumed that there is also a centralized market for contracts with ex ante monitoring. We combine a centralized bond market with price-setting banks to set the stage for the monopolistic banking system assumed in Section 5.
simultaneously announce a deposit interest rate, \( r^d \), and the face value of the loan contract \( R^l(w) \), and commit to transact at these prices with any agent willing to accept the offer. Given \( R^l(w) \) we can easily compute the bank’s expected return, \( r^e(w) \), using Eq. (7). Next, entrepreneurs and workers choose whether to patronize one of the banks or go to the bond market. Bank \( j \) receives \( D_j \) deposits and \( L_j \) loan applications, and both variables are public information. After that, entrepreneurs, workers and banks submit their supplies of and demands for credit to the bond market. In particular, since banks are committed to transact at the announced prices, they will borrow or lend in the bond market \( L_j - D_j \). Finally, transactions in the bond market take place at the implicit market-clearing rate, \( r^b \).\(^{16}\)

In this section we wish to rule out the possibility that any bank may exert market power. Thus, we assume that individual banks face a capacity constraint, which is not binding in equilibrium, but makes sure that each individual bank is small, not only with respect to the banking industry but also with respect to the bond market. As a result banks take \( r^b \) as given. Such a competitive bond market actually separates banks’ decisions in the loan and deposit submarkets.\(^{17}\) Moreover, no bank can behave strategically in one submarket in order to increase its market power in the other by imposing quantity constraints on its rivals.\(^{18}\) Banks unable to capture deposits can still borrow in the bond market and lend to entrepreneurs; and, symmetrically, banks without loan applicants can still lend in the bond market the funds raised from depositors.

Bonds and deposits are perfect substitutes for both banks and workers. Thus, banks will not be able to attract any deposits unless they set \( r^d = r^b \). Finally, if the cost of funds is \( r^b \), then price competition in the loan market implies that \( r^e(w) = r^b \). Consequently, in equilibrium banks make zero profits.

Proposition 4. In equilibrium banks transact with entrepreneurs and workers at the bond rate, i.e. \( r^d = r^e(w) = r^b = r \) and, therefore, banks make zero profits. Moreover, in the limit of equilibria with \( C > 0 \), as \( C \) goes to zero entrepreneur \( i \) trades in the bond market if and only if \( w_i > \frac{r^b}{r} \).\(^{19}\)

Next, we turn to the characterization of the overall equilibrium of the

\(^{16}\) In equilibrium all bonds yield the same rate of return, \( r^b \), although the face value of these bonds, \( R(w) \), will vary with \( w \) according to Eq. (2).

\(^{17}\) In the banking literature this is sometimes called the Monti–Klein theorem.

\(^{18}\) In the absence of capacity constraints, and given the constant returns to scale technology prevailing in the banking industry, competition among intermediaries could give rise to complicated interactions between the deposit and loan submarkets and non-Walrasian outcomes are possible. See, for instance, Stahl (1988) and Yanelle (1989).

\(^{19}\) The statement about \( w_i \) is a direct implication Proposition 3.
model. The bond market clearing determines the relative price of the consumption good in terms of the intermediate good, \( r \); and the number of projects that will be financed, \( w_0 \). Since there is no aggregate uncertainty, rational agents correctly forecast \( r \) and \( w_0 \) when they take their first occupational choice. Hence, the limit of the equilibria of the entire game, as \( C \) goes to zero, is a vector \((s_0, w_0, r)\) such that:

(i) Given \( r \) and \( w_0 \), agents with \( s < s_0 \) choose to become workers (while those with \( s > s_0 \) become entrepreneurs). \( s_0 \) is given by

\[
\frac{r}{2} \int_0^{w_0} f(w|s_0) \, dw + \int_r^{r'} \pi^f(w)f(w|s_0) \, dw + \int_r^{1} \pi^b(w)f(w|s_0) \, dw = r. \tag{7}
\]

(ii) Given \( r \), entrepreneurs with \( w < w_0 \) give up the project at the end of the first stage and use their second half unit of labor in the production of the intermediate good (those with \( w > w_0 \) go on with the project into their second stage). \( w_0 \) is given by

\[
\pi^f(w_0) = r/2. \tag{8}
\]

(iii) The market clears, i.e.

\[
\int_{w_0}^{1} \int_{s_0}^{1} f(w, s) \, ds \, dw = \int_{0}^{1} f(w, s) \, ds \, dw + \frac{1}{2} \int_{0}^{s_0} \int_{0}^{1} f(w, s) \, ds \, dw. \tag{9}
\]

The entrepreneurs’ payoff functions, \( \pi^b(w) \) and \( \pi^f(w) \), are given by Eqs. (4) and (6), respectively, and \( R^e \) is given by Eq. (5). Eq. (7) determines the value of the signal that leaves the agent indifferent between becoming a worker or an entrepreneur. Note that the expected payoff of becoming an entrepreneur, the left-hand side, is increasing in \( s \). Thus, all agents with a higher signal will strictly prefer to develop their investment projects, while agents with a lower signal prefer to become workers.

Similarly, Eq. (8) determines the level of \( w \) that leaves the entrepreneur at the end of the first stage indifferent between quitting or carrying on the project. Again the expected payoff of developing the second stage of the project is increasing in \( w \), so that all entrepreneurs with \( w < w_0 \) will quit and those with \( w > w_0 \) will go on.\(^{20}\)

\(^{20}\) If the occupational choice can only be made after learning \( w \), then an equilibrium is a pair \((w_0, r)\) satisfying conditions analogous to (8) and (9). In this case the system is dichotomous with \( w_0 \) determined exclusively by the market-clearing condition and the private payoff maximizing condition would determine the market return. In this simpler world the role of intermediaries and the entrepreneur’s financial decisions would be similar. The main difference would be that in this case taxes do not distort the allocation of labor.
Sufficient conditions for the existence and uniqueness of an equilibrium are given in the appendix. Next, we compare the competitive equilibrium with the first-best allocation. It is actually easy to see that the first-best allocation is fully characterized by a pair \((s_0, w_0)\) that maximizes aggregate output, \(Y\):

\[
Y = \int_{w_0}^{1} \int_{s_0}^{1} (w + (1/2))f(w, s) \, ds \, dw,
\]

subject to the feasibility constraint (9). \(^{21}\) Proposition 5 shows that, in the absence of government intervention, the competitive equilibrium is inefficient. This result turns out to be crucial to explain the relative efficiency of alternative taxes.

**Proposition 5.** In a competitive equilibrium the levels of consumption, investment and output are inefficiently low, with respect to the first best. As \(D\) goes to zero, the market allocation approaches the ex ante optimal allocation.

The proof can be found in the appendix. The intuition goes as follows. The market return performs two jobs at the same time. On the one hand, it drives the allocation of (labor) resources between the production of the intermediate good and the entrepreneurial activities, and second it determines the frequency of bankruptcies. On the other hand, the lower the return the better (the lower ex post monitoring costs). Unfortunately, the interest rate that clears the market involves a positive frequency of bankruptcies. \(^{22}\)

In the limit case of \(D = 0\), the market allocation is efficient since the frequency of bankruptcies is irrelevant. But as \(D\) increases, an increasing amount of output goes to monitoring and thus consumption falls. Moreover, higher monitoring costs imply a lower return for the marginal entrepreneur, lower credit demand, and lower \(r\). Consequently, \(s_0\) decreases while \(w_0\) increases, and investment also falls.

\(^{21}\) The details can be found in the working paper version; see Caminal (1994).

\(^{22}\) The inefficiency of the competitive equilibrium can be explained in terms of incomplete markets. If agents can trade before they learn their characteristics, then the market allocation would be efficient. However, in this model only spot markets exist, i.e. agents are allowed to trade only after they learn their characteristics.
4. Linear taxes with perfectly competitive banks

In this section we investigate the effect of linear taxes on the competitive equilibrium allocation. Two shortcomings should be pointed out at the outset. First, we constrain ourselves to linear taxes, which may be suboptimal in the presence of fixed costs. Secondly, the statements about optimal taxation are made for the case of an arbitrarily small revenue. Both limitations are imposed by the tractability of the problem.

4.1. A tax on output

Let us consider a proportional tax on output. As shown below, the good news is that such a tax involves no distortion in the labor allocation decision. The bad news is that it involves too much monitoring effort by tax authorities (high tax collection costs).

If we let $\tau$ be the uniform tax rate on output, the analog of Proposition 2 indicates that with a loan contract ex post monitoring occurs if and only if

$$(1 - \tau)(w + x) \leq R^\tau,$$

and $R^\tau$ will be given by

$$R^\tau = \frac{(r/(1 - \tau)) - w}{(1 - \tau)(w + x - D)h(x|a_1) \, dx} + R^\tau \left[ 1 - H \left( \frac{R^\tau}{1 - \tau} - w|a_1 \right) \right] = r. \quad (10)$$

Similarly, the entrepreneur's payoff is given by

$$\pi^\tau(w) = \int_{(r/(1 - \tau)) - w}^{1} \left[ (1 - \tau)(w + x) - R^\tau \right] h(x|a_1) \, dx. \quad (11)$$

Also, entrepreneurs choose to issue bonds if and only if $(1 - \tau)w \geq r$. It is immediate to see that if $r/(1 - \tau)$ is constant with $\tau$, then Eq. (10) implies

$23$ It is assumed that when bankruptcy occurs the tax collection agency bears a proportion $\tau$ of the ex post monitoring costs.
that $R^f/(1 - \tau)$ also remains constant. Therefore, by Eq. (11), $\pi^e(w)/(1 - \tau)$ is also unchanged. Also, note that the set of entrepreneurs that choose to issue bonds remains invariant and $R^b/(1 - \tau)$ as well as $\pi^b(w)/(1 - \tau)$ are also constant. Thus, by checking the equilibrium conditions (7)-(9), we realize that if $r/(1 - \tau)$ is constant with $\tau$, then the allocation of labor is invariant to $\tau$, and so is output. Consequently, $r/(1 - \tau)$ is invariant with $\tau$. Despite the fact that the allocation of resources is unaffected, there is an efficiency loss given that implementing such a tax scheme implies monitoring ex post all projects. This result is summarized in the following proposition:

Proposition 6. A proportional tax on output creates no allocative distortion but involves a fixed collection cost.

The intuition is immediate. The incidence of a tax on output is proportional to all inputs, the intermediate good and entrepreneurial activity, so that no decision is distorted but all expected returns fall in proportion to the tax. However, since monitoring output is costly the government must pay a cost $D$ in all contingencies, except when bankruptcy occurs in which case the government pays $\tau D$. Thus, for small tax rates, revenues will not be able to cover the collection cost.

4.2. A tax on the intermediate good

Let us consider a uniform tax rate $\gamma$ on the intermediate good. We must rewrite Eqs. (7) and (8):

\[
\begin{align*}
(1 - \gamma) \frac{r}{2} \int_0^{w_0} f(w|s_0) \, dw + \int_0^r \pi^e(w)f(w|s_0) \, dw + \int_r^1 \pi^b(w)f(w|s_0) \, dw &= (1 - \gamma)r, \\
\pi^e(w_0) &= (1 - \gamma)(r/2).
\end{align*}
\]

By totally differentiating (4), (6), (9), (12), and (13):

\[
\frac{dr}{d\gamma} > 0, \quad \frac{d(1 - \gamma)r}{d\gamma} < 0, \quad \frac{ds_0}{d\gamma} < 0, \quad \frac{dw_0}{d\gamma} > 0.
\]

The intuition behind the allocative effects of such a tax is the following. Other things equal, a tax on the return from the intermediate good reduces the incentives to become a worker. To eliminate the excess demand for the intermediate good, $r$ must increase, although the after-tax return $(1 - \gamma)r$
falls. A higher \( r \) reduces the expected profits of the marginal entrepreneur, and thus \( w_0 \) increases. However, at the moment of observing the signal, the increase in \( r \) reduces the expected return from entrepreneurial activity but less than the after-tax return on the intermediate good, and hence \( s_0 \) falls.\(^{24}\)

The main idea is that a proportional capital income tax distorts the allocation of resources because only one input is directly taxed. Since the distortion is reflected in the market return on the intermediate good, entrepreneurs also bear part of the tax.

Let us now compute the welfare losses caused by such a tax. We first write the formula for aggregate consumption, \( c \) (which includes private and government consumption):

\[
c = \int_{s_0}^{1} \int_{w_0}^{1} (w + (1/2))f(w, s) \, dw \, ds
- D \int_{s_0}^{1} \int_{w_0}^{1} \int_{0}^{r} Re^{-w} f(w, s)h(x|x_1) \, dx \, dw \, ds.
\]

The first term is aggregate output, while the second term are the aggregate ex post monitoring costs. Using the equilibrium conditions (4), (6), (9), (12) and (13), we can compute the marginal welfare losses of a change in the tax rate:

\[
\frac{dc}{d\gamma} = n_1 \gamma \frac{ds_0}{d\gamma} - n_2 D \frac{dr}{d\gamma},
\]

where \( n_1 \) and \( n_2 \) are positive functions, with exact expressions given at the end of the appendix. The first term is zero for the first dollar (\( \gamma = 0 \)) and negative in general, and it represents the allocative distortion caused by the tax. The second is always negative (even when \( \gamma = 0 \)) and represents the increase in ex post monitoring costs caused by higher interest rates.

Notice that the first dollar collected also has a strictly negative welfare cost. The reason is a pre-existing distortion demonstrated in Proposition 5, i.e. the interest rate in a competitive equilibrium is too high because it induces too many bankruptcies. A tax on the intermediate good raises interest rates and thus increases the frequency of bankruptcies.\(^{25}\) We summarize these results in the following proposition:

\(^{24}\) The market-clearing condition (9) implies that \( s_0 \) and \( w_0 \) must move in opposite directions.
\(^{25}\) If \( D = 0 \), then the welfare costs of the first dollar collected are null.
Proposition 7. A proportional tax on the intermediate good (capital) has strictly positive welfare costs, even for the first dollar collected, since it distorts the allocation of resources and increases the frequency of bankruptcies.

4.3. A tax on bank deposits

Finally, we consider a proportional tax on bank deposits, \( \phi \). Since bank deposits and bonds are perfect substitutes this implies that the before-tax return on deposits is \( \frac{r}{1-\phi} \). Thus, a bank loan must provide a return equal to the before-tax return on deposits, \( \frac{r}{1-\phi} \). Let us rewrite the face value of a loan:

\[
\frac{R^\epsilon - w}{0} (w + x - D)h(x|a_1) \, dx + R^\epsilon \left[ 1 - H(R^\epsilon - w|a_1) \right] = \frac{r}{1-\phi} .
\]

Hence, it is no longer true that the optimal contract involves ex ante monitoring if \( w < r \). The entrepreneur indifferent between applying for a bank loan or issuing bonds is given by the following equation:

\[
\pi^\ell(w_1) = \pi^b(w_1) .
\]

Thus, the equilibrium condition (7) is replaced by

\[
\frac{r}{2} \int_0^{w_0} f(w|s_0) \, dw + \int_{w_0}^{w_1} \pi^\ell(w) f(w|s_0) \, dw + \int_{w_1}^1 \pi^b(w) f(w|s_0) \, dw = r .
\]

Summarizing, the equilibrium conditions are given by (2), (4), (6), (8), (9), and (16)–(18). Totally differentiating these equilibrium conditions:

\[
\frac{dr}{d\phi} < 0, \quad \frac{d(r/(1-\phi))}{d\phi} > 0, \quad \frac{dw_1}{d\phi} < 0, \quad \frac{ds_0}{d\phi} < 0, \quad \frac{dw_0}{d\phi} > 0 .
\]

Most of the intuition is analogous to the previous tax scheme. A tax on bank loans has a different impact on different investment projects, so this explains the distortion in the allocation of labor (the changes in \( w_0 \) and \( s_0 \)). Moreover, a tax on bank loans makes issuing bonds relatively more attractive than applying for a loan, so \( w_1 \) falls with the tax. Finally, since the profitability of entrepreneurial activity falls, \( r \) decreases, although the cost of loans increase (i.e. \( r/(1-\phi) \) increases).

Let us now turn to the welfare loss caused by the tax. We can first rewrite the formula for aggregate consumption:
\[
c = \int_{s_0}^{1} \int_{w_0}^{w} \left( w + f(1, 2) \right) f(w, s) \, dw \, ds
\]

\[
- D \int_{s_0}^{1} \int_{w_0}^{w} f(w, s) h(x | a_1) \, dx \, dw \, ds
\]

\[
- D \int_{s_0}^{1} \int_{w_0}^{w} f(w, s) h(x | a_0) \, dx \, dw \, ds.
\] (19)

The first term is aggregate output, the second term the bankruptcy costs associated with bank loans, and the third the bankruptcy costs associated with bonds. The marginal welfare losses are given by

\[
\frac{dc}{d\phi} = \phi k_1 \frac{ds_0}{d\phi} + \phi k_2 \frac{dw_1}{d\phi} - Dk_3 \frac{d(r/(1 - \phi))}{d\phi},
\] (20)

where \( k_1, k_2 \) and \( k_3 \) are functions (whose exact expressions are given at the end of the appendix), with a positive sign.

The first term captures the distortion in the allocation of labor, which as usual is zero for the first dollar collected. The second term captures the distortion in the firm’s financing decision between bonds and bank loans; again, this distortion is zero for the first dollar of taxes. The third term represents the increase in ex post monitoring costs caused by an increase in the loan interest rate. As in the previous subsection this term is strictly positive even for the first dollar.

Thus, in principle a tax on bank loans creates similar distortions as a tax on the intermediate good but in addition distorts firms’ financial decisions. These results are summarized in the following proposition:

**Proposition 8.** A proportional tax on bank deposits has strictly positive welfare costs, even for the first dollar collected, since it distorts the allocation of resources, the entrepreneurs’ choice between loans and bonds, and increases the frequency of bankruptcies.

We now turn to the issue of comparing the relative efficiency of these alternative forms of taxation. This can be easily done for the first dollar, but unfortunately it becomes quite complicated to keep track of the evolution of the different tax bases as tax rates become strictly positive. The next proposition reports the result for the first dollar of taxes.

**Proposition 9.** With a perfectly competitive banking industry, the most efficient way to collect the first dollar is through a tax on the intermediate
good. In other words, for small tax revenues a tax on the intermediate good strictly dominates the other two forms of taxation.

The proof of this proposition has two parts. (i) Proposition 6 has already shown that the tax on output involves a fixed cost and is therefore unable to raise any positive revenue if the tax rate is small. (ii) From Propositions 7 and 8 we see that the welfare loss of taxation for the first dollar has the same coefficient and only depends on the difference between

$$\frac{dr}{d\gamma} \text{ and } \frac{d(r/(1 - \phi))}{d\phi} \text{ (evaluated at } \gamma = \phi = 0).$$

It can be shown that the second is larger than the first. Moreover, the base of the tax on the intermediate good is larger than the one on bank loans.

It is quite complicated to extend the result to discrete values of the tax rates. However, the insight provided by the analysis suggests that the same result is likely to hold for relatively large tax revenues, since the tax on bank deposits distorts labor allocation in the same direction. It similarly increases the frequency of bankruptcies and, moreover, it distorts the entrepreneurs' financial decisions.\(^{26}\)

5. Taxes with a monopolistic banking industry

In this section we introduce effective market power into banking. The only change in the model is that while all agents still have access to the centralized bond market, the banking industry is now segmented. More specifically, we assume that workers and entrepreneurs are uniformly distributed over a large number of islands. On each island there is a single financial intermediary able to use the ex ante monitoring technology. Agents can hold deposits and apply for loans only at the local bank, but not at banks located on different islands. However, all workers, entrepreneurs and banks can trade in the bond market without any restriction. Thus, banks are local monopolists who compete only with a perfectly competitive bond market.

As in previous sections, banks are assumed to maximize expected profits, which are distributed at the end of the period. For simplicity, the distribution of bank ownership is assumed to be exogenous, and since the

\(^{26}\) One case where the optimal tax scheme changes with the amount of money to be collected is when \(D\) is relatively small. In such a case it could be optimal to collect a substantial amount of revenue through a proportional output tax. The reason is that the distortion caused by alternative taxes increases with the amount collected, but the collection costs of the output tax are fixed.
social welfare criterion is simply aggregate consumption, it need not be specified.

The sequence of moves is exactly as described in Section 4: first, banks set their deposit and loan rates; secondly, agents choose between the bond market and the local bank; and, finally, the bond market clears and transactions take place. In such a framework banks have absolutely no market power in the deposit market, since workers perceive bank deposits and bonds as perfect substitutes. Hence, as in the competitive case, banks' cost of funds is $r^b$. However, they do have market power in the loan market. The reason is that by monitoring entrepreneurs, banks can enforce the efficient action and thus set a higher interest rate.

More specifically, entrepreneurs can only choose between issuing bonds or applying for a loan at the local bank. This implies that the loan interest rate does not induce the entrepreneur to finance its project in the bond market, i.e. $R^e(w)$ must satisfy

$$\int_{R^b-w}^{R^b-w} (w + x - R^b)h(x|a_0) \, dx \leq \int_{R^e-w}^{R^e-w} (w + x - R^e)h(x|a_1) \, dx,$$

(21)

where $R^b$ is given by (2). However, it is immediate to check that the bank has incentives to charge the highest possible interest rate. Therefore, in equilibrium $R^e$ will be determined by Eq. (21) with equality.27

An important remark is that the monopolist can perfectly discriminate across borrowers. This is not an ad hoc assumption but rather the natural consequence of the structure of the model. In fact, this feature is likely to be quite robust to changes in the environment, provided banks perform a role in ex ante monitoring, since they acquire private information about firms' return distribution.

A second remark is that, despite the fact that banks are able to extract all the surplus from firms, the existence of banks is welfare enhancing. The reason is that by monitoring firms' actions they are able to reduce ex post monitoring costs. In fact, banks' profits are equal to the reduction of ex post monitoring costs achieved by ex ante monitoring. Thus, entrepreneurs' expected payoffs are given by

$$\pi^b(w) = w + (1/2) - r, \quad \text{if } w > r,$$

(22)

27 In our model loans are special but bank deposits are not. A more general model of banking would consider explicitly the liquidity services provided by deposits. In this case banks might be able to squeeze some of the depositors' rents and perhaps some of our conclusions on the incidence of different tax schemes would change. In any case, most of the insights provided by our simpler model would probably survive.
\[
\pi^e(w) = w + (1/2) - r - DH(R^b - w|a_0), \quad \text{if } w < r,
\]

(23)

where \(R^b\) is given by (2). Note that the payoff function of a loan applicant, Eq. (23), depends on \(R^b\) because such an entrepreneur is indifferent between loans and bonds. The remaining equilibrium conditions are (7)–(9), as before.

Note that the effect of output and intermediate good taxation would be very similar to the one described in the previous section. However, taxing bank deposits will have qualitatively different effects.

As usual, when taxing banks we need to worry about the determination of \(w_1\). To an entrepreneur with \(w > w_1\) the bank is unable to make positive profits. Thus, the entrepreneur with \(w = w_1\) is indifferent between issuing bonds and applying for a loan and the bank makes zero profits, i.e. \(w_1\) is given by

\[
w_1 + (1/2) - r - DH(R^b - w_1|a_0) = w_1 + (1/2) - \frac{r}{1 - \phi} - DH(R^\ell - w_1|a_1),
\]

(24)

where \(R^\ell\) is given by (21) with equality.

Note that the system is decomposable, and that \(w_0, s_0,\) and \(r\) (and \(R^\ell\)) are independent of \(\phi\), while the only variable affected by the tax is \(w_1\) (and \(R^b\) for \(w_1 < w < r\)). In fact, from Eq. (24):

\[
\frac{dw_1}{d\phi} < 0.
\]

Aggregate consumption is given by Eq. (19), where \(R^\ell\) is implicitly defined by Eq. (21) when it holds with equality. The marginal welfare loss from taxing bank deposits, using Eq. (24), is

\[
\frac{dc}{d\phi} = \frac{dw_1}{d\phi} \frac{\phi r}{1 - \phi} \int_{s_0}^{1} f(w_1, s) \, ds \leq 0.
\]

(25)

Thus, the following proposition holds:

**Proposition 10.** With a monopolistic banking industry the first dollar collected by taxing bank deposits has zero welfare costs. Hence, taxing deposits strictly dominates the other alternative forms of taxation. In fact, the only welfare loss comes from the disintermediation effect.

A tax on bank deposits increases the minimum return that the bank must raise from a loan. Ceteris paribus, the bank will exclude only safe borrowers, i.e. those who can actually borrow from the bond market at rates
that incorporate a low risk premium, while it will keep the rates constant to the rest of the pool of applicants. Since in equilibrium loan applicants are indifferent between issuing bonds or getting a bank loan, disintermediation has no effect on the profitability of entrepreneurial activity. Thus, the allocation of labor is unaffected, and so is the market return. The only effect will be on the amount of ex ante monitoring performed and thus on the bankruptcy costs. The first dollar collected has no welfare cost since the marginal borrower excluded by the bank is one with no risk \((w = r)\), and hence this fact has no effect on ex post monitoring costs. In other words, the tax goes entirely to banks' profits, and since these profits are equal to the amount of bankruptcy costs saved by monitoring entrepreneurs' actions, taxing bank deposits will generally have welfare costs, except for the first dollar.

6. Concluding remarks

In this paper we have analyzed the relative efficiency of taxing financial intermediaries in a stylized general equilibrium model. The optimal tax system is partly determined by the presence of different tax collection costs. These costs are not ad hoc but arise from primitive assumptions about the economy.

A comparison of the results of Sections 4 and 5 (see Table 1) indicates that competitive banks should not be taxed, but it may be optimal to tax

<table>
<thead>
<tr>
<th>Welfare loss through</th>
<th>Distortion in the labor allocation</th>
<th>Increase in the frequency of bankruptcies</th>
<th>Distortion in financial decisions</th>
<th>Tax collection cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax on output</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>yes ((\infty))</td>
</tr>
<tr>
<td>Tax on capital</td>
<td>yes ((0))</td>
<td>yes ((+)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tax on deposits (perfect competition)</td>
<td>yes ((0))</td>
<td>yes ((+)</td>
<td>yes ((0))</td>
<td>-</td>
</tr>
<tr>
<td>Tax on deposits (monopoly)</td>
<td>-</td>
<td>-</td>
<td>yes ((0))</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: Each entry indicates whether the particular tax causes a welfare loss through that particular channel. In the case of yes, it is indicated (between parentheses) whether the distortion is null, positive but finite, or infinite, for the first dollar collected.
banks that enjoy a significant amount of monopoly power. In this respect, our results are in sharp contrast to those of the dominant firm (partial equilibrium) model. In such a framework the optimal scheme consists of taxing the competitive fringe and subsidizing the monopolist. (See, for instance, Mintz and Seade, 1989.)

Clearly, the real world lies somewhere in between the two extreme market structures considered in this paper. The conjecture is that in a more general model of banking (imperfect) competition, the reserve requirement will fall partially on banks, and partially on depositors and loan applicants. But, provided banks' monopoly power is large enough, banks will bear part of the tax and consequently the optimal tax system is likely to involve a positive tax on bank deposits.

The analysis provides some support for the drastic reduction in reserve requirements that has recently taken place in some Southern European countries, in the context of a deregulatory process aimed at enhancing competition in the banking industry. However, it does not automatically justify the high reserve ratios existing in some developing countries, although it does support moderate ratios in those countries with banking sectors characterized by significant degrees of market power.

The empirical evidence available on the incidence of reserve requirements seems to corroborate the view taken in this paper. For instance, Osborne and Zaher (1992) find that the stock prices of large banks are sensitive to announcements of changes in the reserve requirement, which is consistent with the hypothesis that banks bear part of the implicit tax.

The final comment concerns the existence of pure economic profits in banking. It is obvious that a proportional profit tax involves no efficiency loss. In particular, in the model of Section 6, a tax on bank deposits is dominated by a profit tax. One can probably argue that in the real world the observation of banks' economic profits is at least as difficult as the observation of firms' outcomes. However, in the spirit of the paper this implies that a different (and much more complex) model of banking is needed.

All these issues indicate that this paper is only a first step in the right direction, but that clearly further research is needed.

---

28 The reason is that the price is equal to the marginal cost of the competitive fringe but above the marginal cost of the monopolist. Thus, the first dollar collected from competitive firms involves a second-order loss while subsidizing the monopolist reduces the discrepancy between price and marginal cost and involves a first-order gain.

29 Obviously, it would be desirable to perform a similar experiment in intermediate market structures (oligopoly or monopolistic competition). However, modeling bank interaction in an imperfectly competitive framework, when the monitoring role of banks is explicitly recognized, turns out to be quite complicated. See, for instance, Bhattacharya (1992).

30 See also other references cited in that paper.
Appendix

Proof of Proposition 1. We can write the borrower's payoff under this contract, $\pi^b(a)$, as follows:

$$\pi^b(a) = \int_{R^b-w}^{1} (w + x - R^b) h(x|a) \, dx.$$

Thus, the incentives to choose $a_0$ are easily signed:

$$\pi^b(a_0) - \pi^b(a_1) = \int_{R^b-w}^{1} (w + x - R^b)[h(x|a_0) - h(x|a_1)] \, dx$$

$$= (R^b - w)[H(R^b - w|a_0) - H(R^b - w|a_1)]$$

$$- \int_0^{R^b-w} x[h(x|a_0) - h(x|a_1)] \, dx > 0.$$

This sign is positive for $R^b - w < 0.5$, because of Assumption 3. To check this, we define $\phi(z)$ as follows:

$$\phi(z) = z[H(z|a_0) - H(z|a_1)] - \int_0^{z} x[h(x|a_0) - h(x|a_1)] \, dx.$$

Note that $\phi(0) = \phi(1) = 0$, and

$$\phi'(z) = \begin{cases} > 0, & \text{if } 0 < z < 1/2, \\ = 0, & \text{if } z = 1/2, \\ < 0, & \text{if } 1/2 < z < 1. \end{cases}$$

Consequently, $\phi(z) > 0$ for $z \in (0, 1)$.

Proof of Proposition 3. We define $A(w) = \pi^b(w) - \pi^f(w)$, where $\pi^b(w)$ and $\pi^f(w)$ are given by Eqs. (4) and (6), respectively. Then, $A(w) = D[H(R^f - w|a_1) - H(R^b - w|a_0)]$, where $R^f$ and $R^b$ are given by Eqs. (5) and (2), respectively. For $C > 0$, $A(r) > 0$ because the probability of bankruptcy under the bond contract is zero ($R^b = r$). This is not the case with the loan contract, however, because $R^f > r$, since the ex ante monitoring costs have to covered. Also,
Suppose there exists a $\bar{w} \in \{\bar{w}, \ r\}$, such that $A(\bar{w}) = 0$. Then $A'(\bar{w}) > 0$ if and only if $h(R^b - w|a_0) > h(R^e - w|a_1)$, i.e. $R^e - w < x_0$.

Now suppose we let $C$ go to zero. Clearly, $A(r) = 0$ and $A'(r) > 0$ since $h(0|a_1) < h(0|a_0)$. Moreover, using Eqs. (2) and (5) we can check that $R^b(w) > R^e(w)$ for all $w < r$, and hence $A(w) < 0$ for all $w < r$. Finally, $A(w)$ is differentiable and increasing with $C$. Thus, for $C$ small enough there is a unique $\bar{w}$ such that $A(\bar{w}) = 0$, since $\bar{w}$ must be close to $r$, and in this range it has been shown that $A'(\bar{w}) > 0$ whenever $A(\bar{w}) = 0$.

Lemma 1 (Low probability of bankruptcy). Provided $D$ is not too large, and

$$\int_{0}^{1/2} xh(x|a_0) \, dx \geq 1/12,$$

then in the competitive equilibrium:

$$w_0 < r < 1$$

and

$$R^e(w) - w < 1/2 \quad \text{for all } w \geq w_0.$$

An implication of Lemma 1 is that, in equilibrium, there are firms with a positive probability of bankruptcy, but this probability is below 1/2.

Proof of Lemma 1. If $D = 0$, then Eq. (9) becomes:

$$r = (2/3)(w_0 + (1/2)).$$

Since $0 \leq w_0 < 1$, then $w_0 < r < 1$.

With $D = 0$, from Eq. (2) $R^b(w)$ is given by

$$wH(R^b - w|a_0) + \int_{0}^{R^b-w} xh(x|a_0) \, dx + R^b[1 - H(R^b - w|a_0)] = r.$$ 

Note that $R^b$ monotonically decreases with $w$. Is there a $w \geq w_0$ such that $R^b - w = 1/2$? If $R^b - w = 1/2$, then the above expression becomes
Given Eq. (9) and the assumption about the conditional expected value of \( x \) it follows that the value of \( w \) that satisfies the above equation must be below \( w_0 \). Since \( R^g(w) \), \( r \) and \( w_0 \) are continuous in \( D \), provided \( D \) is not too large the inequalities stated above hold.

\[ w + \frac{1}{2} + \int_0^1 xh(x|a_0) \, dx = r. \]

A.1. Existence and uniqueness of equilibria

First, let us introduce some notation:

\[ G(w_0, s_0) = \int_{w_0}^{1} w f(w|s_0) \, dw + w_0 F(w_0|s_0) - \frac{1}{6} - \frac{4w_0}{3}, \]

\[ J(w_0, s_0) = \int_{w_0}^{1} \int_{s_0}^{1} f(w, s) \, ds \, dw - \int_{s_0}^{1} \int_{w_0}^{1} f(w, s) \, ds \, dw \]

\[ - \frac{1}{2} \int_{s_0}^{1} \int_{w_0}^{1} f(w, s) \, ds \, dw, \]

\[ G(\bar{w}_0, 0) = 0, \]

\[ G(0, \bar{s}_0) = 0, \]

\[ J(\bar{w}_0, 0) = 0, \]

\[ J(0, \bar{s}_0) = 0. \]

Lemma 2 (Sufficient conditions). If \( G(0, \bar{s}_0) > 0 \) and \( G(\bar{w}_0, 0) < 0 \), then the competitive equilibrium exists and is unique.

Proof of Lemma 2. The equilibrium conditions (Eqs. (10)–(12)) for \( D = 0 \) can be written as

\[ G(w_0, s_0) = J(w_0, s_0) = 0. \]

It is clear that \( \bar{s}_0 \) and \( \bar{w}_0 \) are strictly less than 1 and that at least one of them is strictly positive. The slope of these two conditions can be easily signed:

\[ \left. \frac{dw_0}{ds_0} \right|_{J=0} < 0, \quad \left. \frac{dw_0}{ds_0} \right|_{G=0} > 0. \]
This implies that if the equilibrium exists, then it is unique. Existence will be guaranteed if

\[ \dot{s}_0 < \ddot{s}_0 \text{ and } \dot{w}_0 < \ddot{w}_0. \]

But since

\[ \frac{\partial G}{\partial s_0}(0, s_0) > 0 \text{ and } \frac{\partial G}{\partial w_0}(w_0, 0) < 0, \]

these conditions are equivalent to \( G(0, \ddot{s}_0) > 0 \) and \( G(\ddot{w}_0, 0) < 0 \). Note also that Eq. (10) is independent of \( D \), Eqs. (8) and (10) are continuous in \( D \), and \( dw_0/\,ds_0 \), conditional on Eqs. (8) and (9), is continuous in \( D \). Hence, if \( D \) is small enough, then there exists a unique equilibrium.

Proof of Proposition 5. The planner's problem consists of maximizing aggregate consumption subject to (9) and subject to the non-negativity constraints \( w_0 \geq 0 \) and \( s_0 \geq 0 \), since it is clear that \( w_0 \) and \( s_0 \) are strictly less than 1. Let us denote by \( \lambda, \delta \) and \( \mu \) the Lagrange multiplier associated with the feasibility and the non-negativity constraints, respectively. The first-order conditions will be given by

\[
\int_{s_0}^{1} (w + (1/2)) f(w|s_0) \, dw = \lambda [2 - (3/2)F(w|s_0)] + \frac{\mu}{f_s(s_0)},
\]

\[
\lambda = (2/3)(w_0 + (1/2)) - \frac{\delta}{\frac{3}{2} \int_{s_0}^{1} f(w_0, s) \, ds}
\]

plus the feasibility constraint. Only one non-negativity constraint can be binding. Suppose, first, that \( \mu > 0 \), i.e. \( \delta = s_0 = 0 \) and \( w_0 > 0 \). Then the first-order conditions imply that \( H(w_0, 0) \). Next, suppose that \( \delta > 0 \), i.e. \( \mu = w_0 = 0 \) and \( s_0 > 0 \). Then the first-order conditions imply that \( H(0, \ddot{s}_0) < 0 \). Finally, if \( \mu = \delta = 0 \), then the first-order conditions imply that \( H(w_0, s_0) = J(w_0, s_0) = 0 \). Consequently, provided the sufficient conditions (given in Lemma 2) for existence and uniqueness of the competitive equilibria are satisfied, the competitive equilibrium with \( D = 0 \) is ex ante efficient.

What if \( D > 0 \)? Totally differentiating (10), (11), and (12) it turns out that

\[ \frac{dr}{dD} < 0, \quad \frac{dw_0}{dD} > 0, \quad \frac{ds_0}{dD} < 0. \]

Since the allocation of labor is distorted from the output-maximizing levels, it follows that with \( D > 0 \), output is lower than in the first best. And
so is consumption: there are lower output costs and higher ex post monitoring costs.

Finally, let us check on the effect of $D$ on the level of investment. Investment, $I$, is given by

$$I = \int_{w_0}^{1} \int_{s_0}^{1} f(w, s) \, ds \, dw.$$

Thus,

$$\frac{dI}{ds_0} = -\int_{w_0}^{1} f(w, s_0) \, dw - \frac{dw_0}{ds_0} \int_{s_0}^{1} f(w_0, s) \, ds$$

$$= \frac{1}{3} \int_{0}^{1} f(w, s_0) \, dw > 0.$$

Thus as $s_0$ falls with $D$, so does investment.

A.2. Coefficients in Eq. (15) in the text

$$n_1 = \frac{r}{3} f_s(s_0),$$

$$n_2 = \int_{s_0}^{1} \int_{w_0}^{r} f(w, s) \frac{h(R^e - w|a_1)}{1 - H(R^e - w|a_1) - Dh(R^e - w|a_1)} \, dw \, ds.$$

A.3. Coefficients in Eq. (20) in the text

$$k_1 = f_s(s_0) \frac{r}{1 - \phi} [\frac{4}{3} - F(w_1|s_0)] > 0,$$

$$k_2 = \frac{r}{1 - \phi} \int_{s_0}^{1} f(w_1, s) \, ds > 0,$$

$$k_3 = \int_{s_0}^{1} \int_{w_1}^{r} f(w, s) \frac{h(R^e - w|a_1)}{1 - H(R^e - w|a_1) - Dh(R^e - w|a_1)} \, dw \, ds$$

$$+ \int_{s_0}^{1} \int_{w_1}^{r} f(w, s) \frac{h(R^e - w|a_0)}{1 - H(R^e - w|a_0) - Dh(R^e - w|a_0)} \, dw \, ds > 0.$$
Acknowledgements

I am grateful to Philippe Bacchetta, Jonathan Hamilton, Kai-Uwe Kühn, Carmen Matutes, Rafael Repullo, Xavier Vives, Marie Odile Yanelle and the anonymous referees of this journal for their useful comments. The research reported in this paper has been financed by the Programa de Estudios Bancarios y Financieros of the Fundación BBV. Partial support from CICYT grants PB/92-0120 and PB/90-0132 is also gratefully acknowledged.

References

Brock, P., 1989, Reserve requirements and the inflation tax, Journal of Money, Credit and Banking 21, 106–121.