Optimal Risk Transfer, Monitored Finance and Real Investment Activity*

Gabriella Chiesa†

Department of Economics
University of Bologna

Abstract

We examine the implications of (optimal) credit risk transfer (CRT) for bank-loan monitoring and financial intermediation. Loans are subject to idiosyncratic risks and to common risk factor. We find that: i) (optimal) CRT enhances loan monitoring and expands financial intermediation, by contrast to the previous literature; ii) optimal CRT's reference asset is loan portfolio; in line with the large development of portfolio products. The intuition is that an optimal contract for the bank to raise finance makes use of the information conveyed by loan-portfolio outcome and rewards the bank as much as possible for the outcomes that signal monitoring: Conditional on monitoring, bank is insulated from exogenous risk (common factor). The amount of capital per lending unit it needs to inject to find it incentive-compatible to monitor attains the minimum; incentive-based lending capacity attains the maximum level. Deposit/debt financing is sub-optimal. It under-rewards monitoring: Bank faces a tighter constraint on outside finance (incentive-based lending capacity is smaller). Optimal CRT addresses these shortcomings: It makes use of the information conveyed by loan portfolio outcome so as to insulate the *monitoring* bank from exogenous risk. Monitoring incentives are enhanced and incentive-based lending capacity attains the maximum. Loan competition is made fiercer, spreads fall, aggregate monitored finance and real investment activity expand. Bank excess return on capital and CRT activity are positively correlated.

Key Words: Credit Risk Transfer, Monitoring Incentives, Competition, Prudential Regulation; JEL classification: G21, D82, G28, D61.

*This work builds on a previous paper that circulated under the title "Risk Transfer, Lending Capacity and Real Investment Activity".

†e-mail: gchiesa@economia.unibo.it

1 Introduction

Traditionally, loans were held on bank balance sheets until maturity or default, and the risk-management tool was the construction of diversified portfolios. That is, there was no discrepancy between real life and the banking paradigm (Diamond, 1984; Ramakrishnan and Thakor, 1984): retaining the loans and diversifying away their idiosyncratic risks allows banks to retain monitoring incentives and hence perform the role of delegated monitors. Moreover, in a Diamond environment of idiosyncratic risks, bank financing with deposits, or more generally with debt, is optimal: bank's reward for monitoring is maximized, and when all idiosyncratic risk is diversified away, the first best solution is attained.

While banks keep raising finance mainly with debt, it is no longer true that credit risk is retained: banks extensively engage in credit risk transfer (CRT). First traded in 1996, the outstanding value of credit derivatives has now reached an amount of about \$4.5 trillion (BIS 2005a). Moreover, the transfer of risk of investment-grade and unrated borrowers, the borrowers that typically need bank monitoring, is steadily increasing (Fitch Ratings, 2004). This phenomenon has attracted the attention of policy makers, national and supranational supervisors, as testified by the huge amount of reports on the topic (e.g. BBA, 2004; BIS, 2003; BIS, 2005a; BIS, 2005b; IMF, 2002). The mixed feelings about the merits of CRT revealed in these reports are well expressed by Warren Buffett arguing that CRT may harm the stability of the financial sector, and Alan Greenspan, opposing the evidence that in the recent US recession corporate failures have neither caused banking failures nor harmed the financial sector. According to Greenspan, this was due to the widespread use by US banks of securitization and credit derivatives. He points out the merits of CRT as a risk-management tool, whereas Buffett points out CRT's (possible) cons of making banks to relinquish their monitoring/screening role.

These developments raise the following important questions. Does CRT harm or rather enhance bank's monitoring, and more generally does it improve welfare? Do banks have incentives to engage in CRT, and what's the role for prudential regulation? What is the implications of a CRT market for banks' strategic interactions in the market place, and hence for bank loan spreads, monitored finance and real investment activity?

The previous literature's result is that CRT harms bank's monitoring incentives, i.e. it undermines the premise of financial stability.[1] However, this negative view has been derived from one-loan bank models and/or by restricting attention to CRT instruments whose reference asset are individual loans. This paper revisits this issue. We impose no (exogenous) restrictions on CRT instruments, and allow for aggregate risk: Loans are subject to idiosyncratic risks and to a common, macroeconomic, risk factor, in line with the evidence that correlation of defaults is driven by the business cycle (BIS, 2005b; Keenan, 2000). We find positive results on CRT: (Optimal) CRT enhances loan monitoring and expands financial intermediation, by contrast to the previous literature. Optimal CRT's reference asset is loan portfolio, in line with the large development of portfolio products. The practical/relevant implication of (optimal) CRT is that the amount of outside finance the bank can raise expands. That is, banks can lever up capital to a greater extent: for any given amount of capital, the lending volume bank finds it incentive-compatible to monitor expands. Or, equivalently, optimal CRT allows banks to economize on capital: any given amount of incentive-compatible lending can be sustained with a lower amount of capital.

The driving force is *monitoring* under *common* risk. Even when all idiosyncratic risk is diversified away, the outcome of loan portfolio is uncertain: For any given bank's monitoring choice, loan portfolio outcome depends on the realization of the common, macroeconomic, factor. Bank my be tempted to bet that exogenous-fortunate circumstances (e.g. the economy's up-turn) will support borrowing firms' performance, avoid costly monitoring and shift the (unmonitored) loans' losses that emerge in the down-turn on to final investors. Bank's moral hazard problem is addressed by two instruments: bank's reward for performance (the carrot), and bank capital per unit of lending (the stick). The higher bank's reward for performance, the lower the amount of capital per lending unit the bank needs to inject to find it incentive compatible to monitor. The higher then bank's outside finance and lending volumes, for any given amount of bank capital. Since bank capital

[1] There is ample evidence that monitoring/screening by banks enhances the quality of financed firms (Datta, Iskandar-Datta and Patel, 1999; James, 1987; Lummer and McConnel, 1989). Moreover, the literature on financial intermediation stresses banks' monitoring role (Campbell and Kracaw, 1980; Diamond, 1984; Fama, 1985; Hellwig, 1991; Bhattacharya and Chiesa, 1995; Holmstrom and Tirole, 1997; and the banking-literature review by Bhattacharya and Thakor, 1993).

is a scarce resource, an optimal contract for the bank to raise finance minimizes the use of the stick and puts more weight on the performance-reward instrument. It makes use of the information conveyed by loan portfolio outcome and rewards the bank as much as possible for the outcomes that signal monitoring: Conditional on monitoring, bank is insulated from exogenous risk (common factor). Debt financing is sub-optimal: It under-rewards bank monitoring. With debt financing, bank faces a tighter constraint on the amount of outside finance it can raise (i.e., is credit rationed). Incentive-based lending capacity is smaller. Optimal CRT addresses the shortcomings of deposit/debt financing. It makes use of the information conveyed by loan portfolio outcome so as to insure the bank against bad luck (bad realizations of the common, macroeconomic, risk factor), if and only if loans are monitored: The amount of capital per lending unit the bank needs to inject to find it incentive-compatible to monitor attains the minimum, bank's incentive-based lending capacity attains the maximum level. Optimal CRT can be implemented by loan portfolio insurance and by loan portfolio securitization.

One result we get is then that, contrary to what is often claimed, CRT enhances banks' monitoring role. Previous literature has shown that CRT weakens monitoring incentives. However, this result has been derived either for one-loan banks (Morrison, 2005; Parlour and Plantin, 2005; Behr and Lee, 2005), and/or by restricting attention to CRT instruments whose reference asset are individual loans (e.g. individual loan sales as in Gordon and Pennacchi, 1995). We find that, in an optimal CRT mechanism the reference asset is loan portfolio; in line with the large development of portfolio products, like, loan securitization, and single CDS (credit default swap) referenced to all the loans (credits) in the portfolio.[2] Second, CRT creates value on incentive-based ground: it lowers the amount of capital a bank must put at stake for finding it incentive-compatible to monitor. It then expands banks' incentive-based lending capacities, and therefore the availability of monitored finance. This differs from the previous literature on CRT where the underlying rationale typically stems from bank risk aver-

[2]The credit derivatives reports by BBA (e.g. Credit Derivatives Reports 2001/02, 2003/04) highlight the rapid increase in portfolio products, and in loan portfolio securitization among these products, which is estimated to account for about 26% of products by the end of 2004.

sion or financial distress costs, or exogenous regulation (Section 1.1 below). Third, (optimal) CRT allows banks to lever up capital to the largest extent. This matches the empirical evidence that banks that engage in CRT have greater leverage and make more business loans (Cebenoyan and Strahan, 2004). Fourth, CRT complements the traditional risk-management tool, diversification: the construction of diversified portfolios allows for (portfolio) CRT instruments that disentangle and transfer exogenous risk. Monitoring incentives are enhanced. Fifth, CRT markets make loan competition fiercer; a by product of CRT's induced expansion of monitored finance availability. Borrowers (firms) are always better off and real investment increases, welfare improves. However, with uncoordinated depositors, for banks to engage in CRT, i.e. for these value/welfare gains to obtain, a condition must hold: capital requirement on loans whose risks are retained (i.e. loans that are not securitized, or more generally, whose risks are not optimally transferred). Indeed, our further results are that, with dispersed depositors/bondholders, banks have incentive to over-loan capital, abstain from CRT, not monitor and shift losses on to lenders (depositors/bondholders). This bank moral-hazard problem is solved by prudential regulation: (incentive-based) capital requirements on loans whose risks are retained. If this condition is satisfied, then banks have incentives to engage in CRT – welfare gains obtain.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 derives the optimal bank's financing contract. Section 4 analyses bank's monitoring incentive problem under debt financing. Section 5 studies monitoring-incentive-compatible CRT mechanisms and derives the optimal one. Section 6 studies bank's incentive to transfer risk and the role that prudential regulation plays. Section 7 derives the implications of CRT for banks' strategic interactions in the loan market, and hence for lending rates, lending volumes, profits, and real investment activity. Section 8 concludes.

1.1 Related Literature

Our analysis is related to various strands of literature. In Duffee and Zhou (2001) deadweight costs associated with bank insolvency provides the incentive to transfer risk. They restrict attention to individual credit risk transfers, credit default swaps (CDS) and individual loan sales, and show that, as long as the asymmetric information about loan quality varies during the life of the loan, credit derivative contracts (CDS) dominate loan sales in circumventing lemon problems caused by banks' superior information about the credit quality of their loans. However, the introduction of credit derivative markets can cause the break down of the loan-sale market. This is detrimental if the asymmetric information problem is one of adverse selection, whereas it may be beneficial if the problem is one of moral hazard, because preventing (individual) loan sales amounts to preventing lower monitoring. We show that CRT in itself enhances monitoring. In Carlstrom and Samolyk (1995) the rationale for CRT relies on ceilings on banks' portfolio risks. In Morrison (2005) credit derivatives accomplish risk sharing by a risk-averse bank. In Wagner and Marsh (2004) difference in banks' and non-banks' bankruptcy costs provides scope for part of the credit risk being transferred to non-banks. Nicolo' and Pelizzon (2005) show how different CRT instruments can be used to signal the quality of bank loans under binding (exogenous) capital requirements.[3] Several authors have analyzed loan sales as an alternative to traditional on-balance-sheet funding that can be cheaper because it signals asset quality (Greenbaum and Thakor, 1987) or because of exogenous reserve and capital requirements (Pennacchi, 1988; Gorton and Pennacchi, 1995). This paper offers a novel explanation for CRT that does not rely on bankruptcy costs, risk aversion, or regulatory constraints. Using an optimal CRT instrument, i.e. transferring exogenous risk, lowers the amount of capital that the bank must put at stake for finding it incentive-compatible to monitor/screen the loans it originates. Monitored finance availability expands, and welfare improves. This paper then complements Arping (2004), who also shows that CRT in itself can enhance banks' monitoring role. In his paper, properly devised risk transfer enhances banks' monitoring role by making banks tougher vis-a-vis poorly performing borrowers (i.e., borrowers' budget constraints harden).

Pennacchi (1988) and Gorton and Pennacchi (1995) analyze a model where banks may improve loan returns by monitoring. Bank regulation (capital and reserve requirements) can give banks an incentive to engage in CRT. They restrict attention to individual loan sales, and derive the optimal individual loan's portion to be sold. This trades off the benefits in terms

[3] De Marzo (2005) shows how assets' pooling and tranching can be used to reduce informational asymmetries.

of regulatory-cost savings, which are increasing in the size of the portion sold, against the costs of under- provision of monitoring, which is higher, the larger is the portion sold. In equilibrium, the bank retains part of the loan, i.e., is exposed to exogenous risk, and chooses a less-than-efficient level of monitoring. A loan sale is a claim on the cash flow from a single loan (i.e. the reference asset is a single loan). By contrast, we have that in an optimal CRT mechanism the reference asset is a loan portfolio, and if the portfolio is sufficiently diversified, then the bank is perfectly insolated from exogenous risk: Monitoring is enhanced and monitored finance availability expands. Morrison (2005) finds the opposite: the possibility of CRT precludes monitoring and leads to disintermediation. In Morrison, a (risk averse) bank deals with one firm, CRT's reference asset is necessarily a single loan. The negative implications of CRT on monitoring incentives are also found by Parlour and Plantin (2005) and Behr and Lee (2005) one-loan bank models.

Froot and Stein (1993, 1998) argue that (otherwise) risk-neutral institutions may well be adverse to cash-flow volatility and engage in risk-management. The argument being that agency problems constrain outside finance and thereby make investment sensitive to inside funds' availability. We show that (optimal) risk-management in itself mitigates/solves agency problems. The availability of outside finance expands.

2 The Model

A bank has the opportunity to extend loans to a continuum of entrepreneurs that can undertake real investment projects but have no resources. Bank has access to project monitoring and screening technologies whose use is costly (see below), it finances lending with its internal funds (capital) and outside finance, this is provided by final investors. The supply of outside finance is perfectly elastic at a gross rate of return which is normalized to one (i.e. the risk free net interest rate is zero). That is, final investors are assumed to make zero profits. Bank acts on behalf of its shareholders, insiders, whose equity holdings constitute bank's endowment of (inside) capital; K is bank's capital endowment.

2.1 Project Technology and Monitoring

A project (firm) requires one unit of resources at date 0, and delivers a return $X \in \{0, x\}$ at date 1. The success probability of each project, i.e. $\Pr(X = x)$, depends on the project type $t \in \{g, b\}$ and on the macro-state realization at the end of period $\theta \in \{\overline{\theta}, \underline{\theta}\}$, where $\overline{\theta}$ denotes the good state realization, the economy's up-turn, and occurs with probability p; $\underline{\theta}$ denotes the bad state realization, the down-turn. If at date 1, the economy is in the up-turn, then a project succeeds for sure no matter its type. If the economy turns out to be in down-turn, then a type g project succeeds with probability $\alpha < 1$, a type b project succeeds with probability $\underline{\alpha} < \alpha$. Whether a bank-funded project is of type g depends on bank's choice of action at the beginning of period $a \in \{m, \varnothing\}$, where m indicates "monitor", and $\varnothing$ "not monitor". Table 1 illustrates project return distribution conditional on bank's action and macro-state realization.

TABLE 1 ABOUT HERE

Monitoring may consist in the provision of services tailored to the firm, or a constraint on the entrepreneur's choice of project through appropriate debt covenants, whose fulfillment is then monitored. This would be the case if, for example, an unconstrained entrepreneur would select the specific project on the basis of the private benefits it can get. Monitoring may also consist in *screening*, i.e. testing of an entrepreneur's credit worthiness (at a cost) in an adverse-selection environment.[4]

Monitoring is costly to the bank: it costs $F > 0$ per project. This is a non-pecuniary effort cost to the bank. F may also be interpreted as the opportunity cost of eschewing insider lending, forgoing the potential private benefits of a type b project in collusion with the borrower.

We specialize the problem by making the following assumptions:

Assumption A1

(an unmonitored project has negative net present value:)

$$\underline{s}x < 1 \quad , \tag{1}$$

[4] In this case, the bank's monitoring cost is the cost of performing a test, divided by the probability that an entrepreneur is endowed with a type g project.

where $\underline{s}$ is the success probability of an unmonitored project:

$$\underline{s} \equiv p + (1-p)\,\underline{\alpha} \ .$$

Assumption A2
(a monitored project has positive net present value:)

$$sx > 1 + F \quad , \tag{2}$$

where s is project success probability conditional upon monitoring:

$$s \equiv p + (1-p)\,\alpha \ .$$

However,

$$(1-p)(1-\alpha x) + F > 0 \ , \tag{3}$$

that is the expected down-side loss of a monitored project, $(1-p)(1-\alpha x)$, plus the monitoring cost, F, is strictly positive. This will imply that, if the bank raises outside finance by issuing debt and abstains from (optimal) credit risk transfer (CRT), then the volume of lending it can make conditional on finding it incentive compatible to monitor, has an upper-bound. Which in turn implies that, in the absence of (optimal) CRT, the bank itself is credit rationed.

Assumption A3
(Bank moral hazard:)
Bank's monitoring/screening choices are unobservable.

Our key assumptions are that: i) monitoring improves loan performance, it's costly and unobservable, i.e. there is bank moral hazard; ii) the average loan return conditional upon bank action is uncertain, it depends on macro-state realization. In other words, credit risk has two components: i) endogenous risk, that which can be controlled by (unobservable-costly) monitoring; ii) exogenous (aggregate) risk, i.e. the common, macroeconomic, risk factor $\theta \in \left\{\overline{\theta}, \underline{\theta}\right\}$.

Let R denote bank's gross lending rate on projects (loans), and let R^*:

$$R^* \equiv \frac{1+F}{s} \ , \tag{4}$$

that is, R^* is the zero-profit lending rate, conditional on monitoring. We assume that bank's lending rate is fixed exogenously at R that satisfies:

$$x \geq R \geq R^* \ , \tag{5}$$

that is bank's lending rate does not exceed project successful outcome, x, and allows to recoup the resources invested in a monitored project (because $R \geq R^*$). In Section 7, we examine how R is determined by competitive considerations.

By assumptions A1-A2, an unmonitored project has negative net present value and a monitored one, positive. Since in equilibrium agents cannot be worse off than with their status-quo payoffs:

Remark 1. *In any equilibrium, bank monitors all loans it originates.*

That is, projects that are financed have positive net present value. Moreover, final investors make zero profits. Then, in equilibrium, bank's profit per unit of lending, π_u, is the surplus generated by a monitored loan:

$$\pi_u = [sR - (1 + F)] \ ,$$

this is positive (by $R \geq R^*$), i.e., bank's profits are increasing in lending. Monitoring incentives entail a constraint on lending. We call this constraint *incentive-based lending capacity*, and proceed as follows. We derive the contract between the bank and final investors (outside-finance providers) that maximizes bank's incentive-based lending capacity, i.e. the optimal contract. This differs from standard debt, which implies that raising deposits, or more generally debt, is suboptimal (i.e. it fails to maximize bank profits). However, there exist CRT instruments such that raising finance with deposits/debt and then engage in CRT, is by all means equivalent to raising finance with the optimal contract: banks' incentive-based capacity is maximized. The CRT instruments that allow for that, i.e. optimal CRT, have as reference asset the loan portfolio.

3 Optimal Contract

An optimal contract between the bank (the agent) and final investors (the principal) will make use of all available information so as to reward the bank "as much as possible" for observable outcomes that signal monitoring. It will then exploit the fact that diversified-portfolio outcomes are centered on their means, and these are conditional on state realization and bank's chosen action. Indeed, consider the loan-solvency rate realization of a diversified portfolio. In state $\overline{\theta}$, loan-solvency rate realization is 1, since in $\overline{\theta}$ all loans perform. In state $\underline{\theta}$, loan-solvency rate realization is α if all loans have been monitored, and lower than α if not loans have been monitored. While a loan-solvency rate realization equal to one does not reveal bank's action, a solvency-rate realization that equals α perfectly reveals that the bank has monitored all loans. Optimal contract rewards the bank "as much as possible" for a solvency rate that equals α, that is for the portfolio outcome that signals that all loans have been monitored. By so doing, the amount of outside finance the bank can raise, conditional on finding it incentive-compatible to monitor, is maximized; that is, bank's incentive-based lending capacity attains its maximum. We show this below.

Let L denote bank's lending and D the amount of outside finance. Because of the flow of funds constraint, $D = L - K$. Consider a contract between the bank and final investors according to which final investors give the bank the amount $D = L - K$ at date 0. At date 1, bank's portfolio revenue accrues to final investors, and these make a transfer payment to the bank contingent on loan-solvency rate realization: W_1, if all loans perform; W_α, if loan-solvency rate realization is α, and $W_{\neq}$, if loan-solvency rate realization is lower than α, which is possible if and only if not all loans are monitored. W_α is bank's reward for monitoring, $W_{\neq}$ is bank's reward for not monitoring all loans. We wish to characterize the portfolio-outcome contingent-payoff schedule, $(W_1, W_\alpha, W_{\neq})$, that maximizes the amount of lending the bank can make, conditional on monitoring being incentive compatible and final investors making zero profits. Clearly, $W_{\neq} = 0$, that is the bank is penalized as much as possible for not monitoring all loans, which also implies that bank's optimal action is either to monitor all loans, $a^* = m$, or monitor no loan, $a^* = \varnothing$. Let bank's payoff schedule be $(W_1, W_\alpha, W_{\neq} = 0)$, then bank's expected profits conditional on $a^* = m$ (monitor all loans), is:

$$\pi(a^* = m) = pW_1 + (1 - p)W_\alpha - FL - K \ ,$$

bank's expected profits conditional on $a^* = \varnothing$ (monitor no loan), is:

$$\pi(a^* = \varnothing) = pW_1 - K$$

The bank monitors if and only if $\pi(a^* = m) \geq \pi(a^* = \varnothing)$, which is true if and only if:

$$W_\alpha \geq \frac{FL}{(1-p)} \tag{6}$$

Condition (6) is the bank's monitoring incentive constraint. The greater the reward for monitoring, W_α, the weaker the constraint. Let condition (6) hold, i.e. bank monitors, then the expected lending revenue is sRL, and final investors make zero profits if:

$$sRL - [pW_1 + (1-p)W_\alpha] = L - K \tag{7}$$

where, the left-hand side of (7) is final investors' expected payoff, conditional on lending being monitored, the right-hand side is the amount of outside finance provided at date 0.

Bank's limited liability requires:

$$W_1 \geq 0 \ \ ; \ \ W_\alpha \geq 0 \ . \tag{8}$$

If bank's payoff schedule $(W_1, W_\alpha, W_{\neq} = 0)$ is optimal, then necessarily it satisfies conditions (6) – (8).

Lemma 1 *Monitoring- incentive compatibility entails a constraint on bank's amount of outside finance and lending; the bigger the payoff bank gets in the up-turn, W_1, the tighter the constraint on (incentive-compatible) lending.*

This follows from the bank's monitoring incentive constraint and the investor's zero profit condition. Indeed, using the final investors' zero-profit condition (7), the monitoring-incentive constraint can be written as:

$$\frac{K}{L} \geq p\frac{W_1}{L} - [sR - (1+F)] \ , \tag{9}$$

that is, the bank finds it incentive compatible to monitor if and only if the amount of capital it invests per lending unit, $\frac{K}{L}$, does not fall below a threshold level, the right-hand side of (9), which is increasing in W_1, the payoff bank gets in the up-turn, i.e. when all loans perform irrespective of monitoring.

Let $W_1 = 0$, then the right-hand side of (9) is negative, because $[sR - (1+F)] > 0$ (by $R \geq R^*$). That is, the bank does not need to inject any capital to find it incentive compatible to monitor: the amount of outside finance the bank can raise, conditional on finding it incentive compatible to monitor, is unlimited: incentive-based lending capacity has no bound. The key is in that a contract that sets $W_1 = 0$ maximizes bank's reward for monitoring, W_α, conditional on final investors making zero profits. For $W_1 = 0$, bank's reward for monitoring is $W_\alpha = \frac{[sRL-(L-K)]}{1-p}$ (by the investor's zero-profit condition (7)). That is, the bank gets the ex-ante expected value of end-of-period wealth conditional on monitoring and final investors making zero profits, i.e. $[sRL - (L - K)]$, multiplied by $\frac{1}{1-p}$, if and only if the loan solvency rate is α, that is, the realization that reveals that monitoring is the chosen action. It gets a zero payoff for any other realization.

However, there is a fundamental problem with a payoff schedule where $W_\alpha > W_1$. Bank's reward for monitoring, W_α, is the payoff bank gets in the down-turn (the state where monitored loans' solvency rate is α); W_1 is the payoff bank gets in the up-turn (the state where all loans perform). Then for $W_\alpha > W_1$, bank's payoff in the down-turn exceeds that in the up-turn, whereas the reverse holds for portfolio performance. The contract is then vulnerable to portfolio outcome falsification: in the up-turn, the bank profits by forgiving part of firms' debt so as to mimic the performance of a monitored portfolio in the down-turn. By so doing it gets the high payoff W_α instead of the nil payoff W_1. Falsification then imposes a further constraint on the bank's payoff schedule, the monotonicity constraint:

$$W_\alpha \leq W_1 \ . \tag{10}$$

The monotonicity/feasibility constraint imposes the upper bound W_1 on the bank's monitoring reward W_α. Optimal payoff schedule maximizes bank's reward for monitoring subject to investor's zero-profit condition (7), and the monotonicity/feasibility constraint (10). Then, denoting optimal values with stars, in an optimal payoff schedule:

$$W_\alpha^* = W_1^* \equiv sRL - (L - K) \quad , \tag{11}$$

and the bank monitors if and only if:

$$\frac{K}{L} \geq c^*$$

$$c^* \equiv -(sR - 1) + \frac{F}{1-p} \ . \tag{12}$$

c^* is the minimum amount of capital per lending unit the bank needs to inject to find it incentive compatible to monitor. Bank's incentive-based lending capacity is $\mathcal{L}^*$:

$$\mathcal{L}^* = \begin{cases} \frac{K}{c^*} & \text{if } c^* > 0 \\ \\ \infty & \text{if } c^* \leq 0 \end{cases} \tag{13}$$

If $c^* \leq 0$, then the incentive-based lending capacity has no bound: the amount of outside finance the bank can raise is unlimited. If $c^* > 0$, then bank's outside finance is constrained: it cannot exceed $\frac{(1-c^*)}{c^*}K$.

For $L \leq \mathcal{L}^*$, bank monitors: Its end-of-period wealth in the down turn amounts to W_α, and this equals bank's terminal wealth in the up-turn, W_1; bank's end of period wealth is the same in every state. The same insurance result holds for profits, since these obtain by subtracting capital (that is bank's initial wealth) and monitoring costs from end-of-period wealth .

The foregoing can be summarized in the following proposition.

Proposition 1 *An optimal contract makes use of the information that loan portfolio's return conveys on bank's monitoring, and maximizes bank's reward for monitoring subject to the investor's zero-profit constraint and the monotonicity constraint on bank's payoff schedule: i) the minimum amount of capital per lending unit the bank needs to invest to find it incentive-compatible to monitor attains its minimum level* c^*, *incentive-based lending capacity attains its maximum* $\mathcal{L}^*$; *ii) conditional on lending not exceeding capacity, bank monitors and is insured against exogenous risk: bank profits are the same in every state.*

4 Debt Financing

For reasons that are outside this paper's scope, like liquidity services' provision and tax benefits, banks raise finance with deposits and more generally by issuing debt. We study bank's incentive problems under debt financing and show that monitoring incentives entail tighter constraints on outside finance (leverage) and on the lending volume that can be made for any given level of bank capital.

Let bank's lending volume be L, assume that its gross borrowing rate is equal to one (which is always true in equilibrium), then bank's contractual debt repayment, due at the end of the period, equals the amount of debt issued, that is $L - K$, and:

bank's end-of-period wealth in the up-turn (i.e., state $\overline{\theta}$) is $W_{\overline{\theta}}$:

$$W_{\overline{\theta}} = RL - (L - K) \ ,$$

bank's end-of-period (pecuniary) wealth in the down-turn, conditional on monitoring, is $W(\underline{\theta}|m)$:

$$W(\underline{\theta}|m) = \max\left[\alpha RL - (L - K), 0\right] \ ,$$

where $W(\underline{\theta}|m) = 0$ if the bank is insolvent, i.e. if debt's repayment, $L - K$, exceeds the outcome of monitored-loan portfolio in the down-turn, αRL.

With debt, the bank's payoff schedule is a portfolio-outcome-contingent payoff schedule, $(W_1, W_\alpha, W_{\neq})$, that sets $W_1 = W_{\overline{\theta}}$, and $W_\alpha = W(\underline{\theta}|m)$. Clearly $W_\alpha < W_1$. With debt financing, bank's monitoring is under-rewarded: incentive-based lending capacity shrinks (by Lemma 1).

Define c^D

$$c^D \equiv (1 - \alpha R) + \frac{F}{(1 - p)} \qquad . \tag{14}$$

Note that $c^D > c^*$ (by (12)). We prove below that c^D is the minimum amount of capital per lending unit that a debt-financed bank needs to invest to find it incentive-compatible to monitor.

Proposition 2 *Let the bank raise finance with debt. Let it borrow at a gross rate equal to one, then if lending volume satisfies:*

$$\frac{K}{L} \geq c^D \qquad ,$$

its profit-maximizing choice is to monitor all loans, it is solvent with probability one, its expected profits are non-negative, and its lenders make zero profits. By contrast, if $\frac{K}{L} < c^D$, then for any borrowing rate, the bank earning non-negative expected profits would necessarily imply that its profit-maximizing choice is to monitor no loan, and its lenders suffer expected losses.

The key observations are that the bank has limited liability, a monitored lending portfolio performs better than an unmonitored one in the down–turn (i.e., state $\underline{\theta}$), and a monitored (unmonitored) loan has positive (negative) net present value. Then, partial monitoring, that is monitoring a fraction $0 < \lambda < 1$ of the loan portfolio, is a strictly dominated strategy. Indeed, if the bank chooses $0 < \lambda < 1$ and is insolvent in $\underline{\theta}$, then it would be better off by choosing $\lambda = 0$, that is monitor no loan, thus avoiding monitoring costs altogether. By contrast, if it is solvent, then it suffers all the consequences of financing $(1 - \lambda) L$ projects with negative net present value, so it would be better off choosing $\lambda = 1$, that is monitor all loans. Bank's undominated strategy is then either to monitor all loans, $a^* = m$, or monitor no loan, $a^* = \varnothing$. And:

(i) If bank's profit-maximizing choice is to monitor all loans (that is, if $a^* = m$), then it is necessarily true that the bank is solvent in the down-turn, that is, it enjoys the benefits of its costly monitoring; which also implies that the bank is solvent with probability one and its borrowing rate is equal to one. Bank's expected profits, conditional on monitoring being the profit-maximizing choice, are then:

$$\pi(a^* = m) = pW_{\overline{\theta}} + (1 - p) [\alpha RL - (L - K)] - FL - K \tag{15}$$

(ii) If bank's profit-maximizing choice is not to monitor (that is, if $a^* = \varnothing$), then it is necessarily the case that its lending volume is sufficiently large with respect to capital to ensure that the bank is insolvent in $\underline{\theta}$ (if it were solvent then it would suffer all the consequences of financing projects with negative net present value, and its profit-maximizing choice would necessarily be to monitor). Hence, if $a^* = \varnothing$, then bank is solvent *only if* the macro-state realization is good $(\overline{\theta})$, when loans perform well regardless of whether they are monitored. In state $\underline{\theta}$, bank's end-of-period wealth is nil, so its expected profits are:

$$\pi(a^* = \varnothing) = pW_{\overline{\theta}} - K \tag{16}$$

Bank's profit-maximizing choice is to monitor if and only if:

$$\pi(a^* = m) \geq \pi(a^* = \varnothing) \ ,$$

that is, by using (15) – (16) :

$$\alpha RL - (L - K) \geq \frac{FL}{(1 - p)} \ , \tag{17}$$

which holds if and only if:

$$\frac{K}{L} \geq c^D \ .$$

Inequality (17) is the bank's monitoring incentive constraint under debt financing. The left-hand side is the monitoring reward of a debt-financed bank. For any given amount of outside finance, $L - K$, monitoring reward under debt financing is lower than that provided by an optimal contract, $\alpha RL - (L - K) < W_\alpha^* \equiv sRL - (L - K)$ (by (11)). The amount of capital per lending unit the bank needs to inject to find it incentive compatible to monitor expands, $c^D > c^*$ (by (12), (14)). Incentive-based lending capacity shrinks to $\mathcal{L}^D$:

$$\mathcal{L}^D = \frac{K}{c^D} \quad , \tag{18}$$

which in turn implies the upper-bound constraint, $\frac{(1-c^D)}{c^D}K$, on the amount of debt that the bank can raise conditional on its lenders making non-negative profits. Section 6 discusses the policy implications that arise when finance is provided by dispersed (uncoordinated) final investors, e.g. depositors.

The lower the monitored loan's down-side loss, i.e. $(1 - \alpha R)$, the lower then the minimum (incentive-compatible) capital requirement, c^D, and hence the greater incentive-based capacity, for any given level of bank capital K: Transferring *monitored* loans' down-side losses expands bank's reward for monitoring and thereby expands incentive-based capacity.

5 Optimal CRT

We noted above that transferring *monitored* loans' down-side losses expands lending capacity. This provides the scope for CRT (credit risk transfer). This section studies incentive-compatible CRT mechanisms and derives the optimal one. The key results are that: i) in an optimal CRT mechanism, the reference asset is loan portfolio; ii) debt financing cum optimal CRT implements the solution attained by an optimal contract: bank's incentive-based lending capacity attains its maximum $\mathcal{L}^*$. And, conditional on lending not exceeding capacity, the bank is insured against exogenous risk: bank profits are the same in every state, iii) optimal CRT can be implemented by loan-portfolio securitization.

Incentive-compatible CRT instruments disentangle and transfer solely exogenous risk: insure the bank against bad luck (down-turn realization) conditional on monitoring. An optimal arrangement will then make use of all available information to infer whether the bank has monitored. The information content of portfolio outcome dominates that of individual loans' outcomes: An optimal CRT deal necessarily belongs to the class of portfolio CRT mechanisms: the reference asset is the loan portfolio and the credit event is the (contractually defined) portfolio's default event.

5.1 Portfolio Risk Transfer

Consider a diversified loan portfolio of size L. Then, in the down-turn, the amount of non-performing loans is $(1-\alpha)L$, if all loans have been monitored, and is higher than that if not all loans have been monitored. In the up-turn, all loans perform, irrespective of monitoring.

Consider a portfolio CRT contract, (P, Y), according to which: if non-performing loans amount to $(1-\alpha)\,L$, then the insurer pays Y to the bank, and if non-performing loans amount to anything else, then bank pays P to the insurer (subject to bank's limited liability).[5] The key feature of this contract is in that the bank is indemnified against loan losses if and only if non-performing loans amount to $(1-\alpha)\,L$, that is if and only if loans have been monitored.

Let bank's lending be L, financed partly with capital and the remaining, $L-K$, with debt. Let bank's gross borrowing rate be equal to one, which is always true in equilibrium (by the same reasoning as in Proposition 2). Let the bank enter into a CRT contract, (P, Y), where $Y > 0$. Then, by the same reasoning as in Section 4, the bank's monitoring-incentive constraint is:

$$\alpha RL - (L-K) + Y \geq \frac{FL}{(1-p)} \;, \tag{19}$$

which holds if and only if:

$$\frac{K}{L} \geq c$$

$$c \equiv \left(1 - \alpha R - \frac{Y}{L}\right) + \frac{F}{1-p} \tag{20}$$

[5] The (P, Y) contract is by all means equivalent to a deal where i pays P up front and gets from the insurer the amount $P + Y$ if and only if $(1-\alpha)\,L$ loans default.

Portfolio CRT expands bank's reward for monitoring, i.e. the payoff bank gets in the down-turn conditional on monitoring. Indeed, conditional on monitoring, non-performing loans amount to $(1-\alpha)\,L$ in the down-turn: bank pockets insurer's transfer-payment Y. Portfolio CRT then relaxes the monitoring incentive constraint (by comparing (19) with (17)): the minimum amount of capital per lending unit that the bank needs to inject to find it incentive compatible to monitor shrinks (by comparing (20) with (14)) – incentive-based lending capacity expands. The larger transfer-payment Y, the more the bank is rewarded for monitoring, the weaker the monitoring incentive constraint (19), and the lower c, i.e. the minimum amount of capital per lending unit the bank needs to inject to find it incentive compatible to monitor. Optimal CRT is then (P^{*}, Y^{*}):

$$(P^{*}, Y^{*}) = \arg\max_{(P,Y)} Y \tag{21}$$

$$s.t.$$

$$pP - (1-p)Y = 0\ , \tag{21.a}$$

$$\alpha RL - (L-K) + Y \leq RL - (L-K) - P \tag{21.b}$$

Inequality $(21.a)$ is the insurer's zero-profit constraint, conditional on loans being monitored (which is true for $\frac{K}{L} \geq c$): the insurer pays Y to the bank in the down turn, and gets bank's payment P in the up-turn. Condition $(21.b)$ is the monotonicity constraint on the bank's payoff schedule (already discussed in Section 3). This constrains (P, Y) to be such that, conditional on monitoring, bank's end-of-period wealth in the down-turn, i.e. the left-hand side of $(21.b)$, does not exceed bank's end-of-period wealth in the up-turn. If $(21.b)$ would fail to hold then, in the up-turn, the bank would profit by forgiving part of firm debts so as to mimic monitored portfolio's performance in the down-turn and thereby pocket insurer's transfer payment Y. The monotonicity/feasibility constraint sets an upper-bound on Y, i.e. $Y \leq (1-\alpha)\,RL - P$, which then implies, by using the insurer's zero-profit condition $(21.a)$, that:

$$Y^{*} = p\,(1-\alpha)\,RL\ ,$$

$$P^{*} = (1-p)\,(1-\alpha)\,RL\ \ .$$

Bank's reward to monitoring is now $(\alpha RL - (L-K) + Y^*)$, and the monitoring-incentive constraint (19), i.e.

$$(\alpha RL - (L-K) + Y^*) \geq \frac{FL}{(1-p)} \quad ,$$

holds if and only if:

$$\frac{K}{L} \geq c_{RT}$$

$$c^{RT} \equiv c^* \equiv -(sR_i - 1) + \frac{F}{1-p}$$

c^{RT} is the minimum amount of capital per unit of lending that a debt-financed bank needs to invest to find it incentive-compatible to monitor, under optimal CRT. It is identical to that attained by an optimal financing instrument. The incentive-based lending capacity of a debt-financed bank that engages in optimal CRT, $\mathcal{L}^{RT}$, is then $\mathcal{L}^{RT} \equiv \mathcal{L}^*$. That is, optimal CRT, maximizes incentive-based capacity.

Conditional on lending not exceeding capacity, bank monitors and its end-of-period wealth in the down-turn is $W_{\underline{\theta}}$:

$$W_{\underline{\theta}} = [\alpha RL - (L-K)] + Y^*$$

$$\equiv sRL - (L-K) \ ,$$

this is identical to $W_{\overline{\theta}}$, bank's end-of-period wealth in the up-turn:

$$W_{\overline{\theta}} = [RL - (L-K)] - P^*$$

$$\equiv sRL - (L-K) \ .$$

Proposition 3 *In an optimal CRT mechanism: i) the reference asset is a loan portfolio; ii) debt financing cum optimal CRT implements the solution attained by an optimal contract: the amount of capital per lending unit that the bank needs to invest to find it incentive-compatible to monitor attains its minimum level* $c^{RT} \equiv c^*$*, incentive-based lending capacity attains its maximum* $\mathcal{L}^{RT} \equiv \mathcal{L}^*$*. Conditional on lending not exceeding capacity, the bank monitors and is insured against exogenous risk: bank profits are the same in every state.*

In an optimal CRT scheme, the bank retains endogenous risk and is insulated from exogenous risk: incentive-based capital requirement shrinks and thereby lending capacity expands. We prove below that loan-portfolio securitization is an optimal CRT scheme.

5.2 Loan-portfolio Securitization

The reference asset is a loan portfolio and profits of the loans' originator (the bank) are the same in every state (i.e. exogenous risk is transferred). We examine equilibrium price for loan-portfolio-backed securities and show that loan securitization implements optimal CRT.

Let the bank securitize the loan portfolio: it sells L securities, each paying $\frac{1}{L}$ of the loan-portfolio return at the final date.

Define:

condition (i): bank's lending does not exceed incentive-based capacity $\mathcal{L}^{RT}$:

$$L \leq \mathcal{L}^{RT} \quad ;$$

condition (ii): security holders are given the option to sell securities back to the bank (the loan originator) at unit price G:

$$G \equiv \alpha R \quad .$$

Let securitization satisfy condition (ii). At the final date, if the state realization is down-turn: if the bank has monitored, then the security market price is $\alpha R \equiv G$, options are not exercised; if the bank has not monitored, then the security market price is lower than αR, options are exercised.

Lemma 2 *If bank's loan-portfolio securitization satisfies conditions (i)- (ii), then the price bank gets per security is $S^m \equiv sR$, and the bank monitors. If either i) and/or ii) fails to hold, then the price bank gets per security is $S^{\varnothing} \equiv \underline{s}R$, and the bank does not monitor.*

Proof: See Appendix B ■

The key is that monitoring incentives rely on two ingredients: (i) the amount of bank's lending relative to capacity; and (ii) the put options security holders are provided with. The put option feature acts as a (off-equilibrium) threat: if the bank does not monitor, then security holders will exercise their put options and as a result, the bank will be asked to pay $R(\alpha - \underline{\alpha})L$ in the down-turn, the state where monitoring upgrades firms' probability of success from $\underline{\alpha}$ to α. However, the bank is protected by limited liability, and if (i) fails to hold, that is if $L > \mathcal{L}^{RT}$, then it finds it optimal to forego costly monitoring altogether, pocket the gains in the up-turn and default on its obligations in the down-turn. Condition (i), that is, lending not exceeding capacity, is then necessary to provide incentives. Loan-portfolio securitization that satisfies (i) and that gives security holders put options at (ii) makes the bank retain endogenous risk, the risk which is transferred is solely the exogenous one. The bank monitors, the expected payoff of a security is the expected return of a monitored loan, and this equals the security price S^m – i.e. the price that market participants are willing to pay for securities that are backed by monitored loans.

Proposition 4 *Loan-portfolio's securitization cum put options is an optimal CRT scheme: the amount of capital per lending unit that the bank must pledge to find it incentive-compatible to monitor attains its minimum level $c^{RT} \equiv c^*$, incentive-based lending capacity attains its maximum $\mathcal{L}^{RT} \equiv \mathcal{L}^*$.*

Proof: By Lemma 2 and Proposition 3 ■

If securitization meets conditions (i)-(ii), then security price is $S^m \equiv sR$, and the bank monitors (by Lemma 2). Securitization cum put options then allows the bank to lend up to the maximum incentive-compatible level, $\mathcal{L}^{RT} \equiv \mathcal{L}^*$, and to earn (monitored lending) profits $\pi(a^* = m) \equiv [sR - (1 + F)]L$. The solution is the same as that attained by (P^*, Y^*), the portfolio risk-transfer scheme which is optimal given the no-falsification (monotonicity) constraint. The bank retains endogenous risk and is insulated from exogenous risk. This would not be true for a partial loan sale (as in Gorton and Pennacchi, 1995; Pennacchi, 1998). In a partial loan sale, or equivalently a loan sale with partial recourse, the bank is necessarily exposed to exogenous risk: the amount of capital per lending unit it needs to pledge is higher (incentive-based capacity is smaller).

6 Bank's Incentive to Transfer Risk: the Role of Prudential Regulation

Is there a need for prudential regulation? We show below that the answer is positive, although not obvious. Indeed:

Lemma 3 *If the bank engages in CRT, then necessarily it monitors, $L \leq \mathcal{L}^{RT}$, and its profits are $\pi(a^* = m)$.*

The key is that a loan-portfolio securitization that satisfies conditions (i)-(ii) perfectly reveals that transferred risk is monitored risk (by Lemma 2). That is:

a) if bank's lending does not exceed $\mathcal{L}^{RT}$, and loans are securitized cum put options, then the price it gets per security is S^m:

$$S^m \equiv sR \geq 1 + F \quad (\text{by } R \geq R^*) ,$$

bank monitors all loans, repays its lenders (depositors/bondholders) and its profits are $\pi(a^* = m) \equiv [sR - (1 + F)] L$, which is positive (by $R \geq R^*$);

b) if it has lent in excess of $\mathcal{L}^{RT}$ and securitizes the loans, then the price it gets per security is $S^{\varnothing}$:

$$S^{\varnothing} \equiv \underline{s}R < 1 \quad (\text{by assumption A1}) ,$$

it fails for sure – its profits are strictly negative (because K is lost).[6]

Therefore, if the bank engages in CRT then necessarily it finds it optimal to monitor: $L \leq \mathcal{L}^{RT}$, and its profits are $\pi(a^* = m)$.

What prevents banks to over-loan capital, i.e. lend in excess of incentive-based capacity $\mathcal{L}^D$, abstain from CRT, not monitor and shift resulting losses on to dispersed depositors? Not dispersed depositors, because these cannot constrain bank's loanable funds. And certainly not bank shareholders, since these profit by over-loaning capital: $\pi(a^* = \varnothing) > \pi(a^* = m)$, for $\forall\ L > \mathcal{L}^D$ (by (17)). It's capital requirements on loans whose risks are retained,

[6] Similarly, if the bank takes the portfolio risk transfer scheme (P, Y), then for $L > \mathcal{L}^{RT}$ it never gets the transfer payment Y, because for $L > \mathcal{L}^{RT}$ the bank does not monitor: non-performing loans exceed $(1 - \alpha) L$. For $L > \mathcal{L}^{RT}$, the risk transfer scheme (P, Y) such that the insurer makes zero profits then sets $P = 0$, i.e. for $L > \mathcal{L}^{RT}$, CRT does not occur.

i.e. loans that are not securitized, or more generally whose risks are not (optimally) transferred, that can effectively prevent banks from over-loaning capital and gain at depositors' expense.

Proposition 5 *i) If bank engages in CRT, then necessarily it monitors:* $L \leq \mathcal{L}^{RT}$ *; ii) bank engages in CRT, if capital requirement on loans whose risks are retained does not fall below the incentive-based threshold* c^D*; iii) If prudential regulation fails to satisfy condition ii), then (uncoordinated) final investors do not fund banks: intermediation fails to occur.*

If prudential regulation on retained risks fails to satisfy condition ii), that is, if capital requirement on loans that are not securitized, or more generally whose risks are not (optimally) transferred, falls below the incentive-based threshold c^D, then the bank finds it optimal to lend in excess of incentive-based capacity $\mathcal{L}^D$, it abstains from CRT, and does not monitor, because $\pi(a^* = \varnothing) > \pi(a^* = m)$, for $\forall\ L > \mathcal{L}^D$. Final investors will rationally anticipate that and will therefore refuse to fund banks, with the exception of the unlikely environment where they can coordinate their decisions so as to set contractually, and enforce, bank's capital requirement on retained risks at the monitoring-incentive compatible level. By contrast, if capital requirement on retained risks does not fall below c^D, then the bank never lends in excess of $\mathcal{L}^{RT}$, and it engages in CRT for $L > \mathcal{L}^D$. That is, the bank monitors and depositors do not make losses. Capital requirement on retained risks does not constrain lending; it supports the monitoring risk-transfer equilibrium.

7 CRT, Loan Competition, and Real Investment Activity

We derive the effects of CRT on banks' strategic interactions in the lending market, and hence on lending rates, lending volumes, bank profits and real investment activity. To discuss this, we allow for n banks (indexed by $i = 1, 2..., n$), and consider a rectangular loan demand curve: there is a measure M of entrepreneurs, each entrepreneur has the real investment opportunity described in Section 2.1. At date 0, each bank $i = 1, 2..., n$, offers a gross rate R_i (possibly, the outcome of a randomization strategy) at which is willing to lend. Each entrepreneur ranks banks according to their lending rates, applies for a loan to his most preferred bank (the lowest-rate-setting bank), if denied credit, which is possible, because banks face incentive-based capacity constraints, applies to his second-preferred bank,...and eventually applies to the highest-rate setting bank if he did not obtain a loan by any other bank. The demand for loans of bank i is then a function of its lending rate, R_i, and competitors' rates, R_{-i}, and is denoted $L_i^d(R_i, R_{-i})$. The lowest-rate-setting bank's loan demand is M (because all entrepreneurs are willing to borrow at the lowest rate), the loan demand faced by the second-lowest-rate-setting bank is the left-over of the lowest-rate-setting bank, and is smaller the larger the lending capacity of the latter. The highest-rate-setting bank's loan demand is the left-over of all the other banks. Bank $i\prime s$ lending volume is the minimum between the demand it faces and its lending capacity. Lending is financed with capital and debt raised at the (equilibrium) gross rate equal to one.

Let c denote the capital requirement on loans whose risks are retained, and let c:

$$c \geq (1 - \alpha R_i) + \frac{F}{1-p} \quad , \quad \forall i \quad , \tag{22}$$

that is,

$$\frac{K_i}{c} \leq \mathcal{L}_i^D \ , \ \forall i \ ;$$

which ensures financial intermediation (by Proposition 5).

Given bank's option to engage in credit-risk transfer (CRT), for any given $i's$ lending rate, R_i, bank $i's$ lending capacity is $\mathcal{L}_i$:

$$\mathcal{L}_i \in \left\{ \frac{K_i}{c}, \mathcal{L}_i^{RT} \right\}$$

where $\frac{K_i}{c}$ is the maximum lending volume that i can make, conditional on abstaining from CRT (retain risks); $\mathcal{L}_i^{RT}$ is the maximum lending volume that i finds it optimal to make, conditional on engaging in CRT (by Lemma 3).

Clearly:

$$\mathcal{L}_i^{RT} > \frac{K_i}{c} \quad ,$$

because $\mathcal{L}_i^{RT} > \mathcal{L}_i^D \geq \frac{K_i}{c}$ (by (22)).

For any given R_i and competitors' lending rates, R_{-i}, the demand for lending faced by i is $L_i^d\left(R_i, R_{-i}\right)$, and $i's$ lending volume is $L_i\left(R_i, R_{-i}; \mathcal{L}_i\right)$:

$$L_i\left(R_i, R_{-i}; \mathcal{L}_i\right) = \min\left[L_i^d\left(R_i, R_{-i}\right), \mathcal{L}_i\right] \ ,$$

$i's$ profits are $\pi_i\left(R_i, R_{-i}; \mathcal{L}_i\right)$:

$$\pi_i\left(R_i, R_{-i}; \mathcal{L}_i\right) = \left[sR_i - (1+F)\right] L_i\left(R_i, R_{-i}; \mathcal{L}_i\right) \quad ,$$

where $\left[sR_i - (1+F)\right]$ is the profit per lending unit – i.e. the profit per monitored loan (by $L_i(..) \leq \mathcal{L}_i$). This is positive, for any lending rate $R_i \geq R^*$. Therefore:

$$\mathcal{L}_i^{RT} = \underset{\mathcal{L}_i \in \left\{\frac{K_i}{c}, \mathcal{L}_i^{RT}\right\}}{\arg\max} \quad \pi_i\left(R_i, R_{-i}; \mathcal{L}_i\right)$$

Lemma 4 *Bank's individually-rational profit-maximizing choice is to self-restrain its lending neither above nor below the incentive-based capacity that optimal CRT allows for:* $\mathcal{L}_i = \mathcal{L}_i^{RT}$, $i = 1, ..n$.

CRT opportunities thus weaken the restraints on bank lending volumes: banks compete more aggressively, borrowers are better off.

By the same reasoning as in Chiesa (2001), the equilibrium of loan competition depends on structural parameters, that is, bank capital, K_i, $i = 1, ...n$, aggregated demand for loans, M, and project return distribution. If aggregate bank capital is sufficiently scarce, that is if:

$$\sum \mathcal{L}_i^{RT} \leq M \ , \text{ for } R_i = x \ , \ i = 1, ..n \ , \tag{23}$$

or, equivalently

$$\left[\frac{1}{-(sx-1) + \frac{F}{(1-p)}}\right] \sum K_i \leq M \ ,$$

then the overall amount of lending that banks can make does not exceed aggregate demand: the equilibrium strategy of a bank is to offer the monopoly rate x. Credit will be rationed if inequality (23) is strict.

If there are at least two banks, i and j, that are sufficiently capitalized so as to finance all firms at the zero profit rate, that is, if :

$$\begin{array}{c} \mathcal{L}_i^{RT} \geq M \ , \ \text{ for } R_i = R^* \\ \mathcal{L}_j^{RT} \geq M \ \ , \text{ for } R_j = R^* \ , \ \ j \neq i \end{array} \quad , \tag{24}$$

then a bank's competitor can cover the whole market, competition is unrestricted and in equilibrium bank profits are driven to zero. All firms borrow at the zero-profit rate, R^*, and the first-best optimum is attained.

If neither condition (23) nor (24) holds, then there is no equilibrium in pure strategies, for the same reasons as in the standard Bertrand-Edgeworth model. In equilibrium, each bank $i = 1, .., n$, randomizes its lending rate according to a distribution function with (common) support $[\underline{R}, x]$, where the lower bound $\underline{R}$ exceeds the zero profit rate, R^*. That is, lending rates exceed the zero-profit value, bank profits are strictly positive, albeit below the monopoly level. The higher bank lending ceilings, the smaller the amount of lending by a bank when it is undercut by competitors, and when it sets its rate it places more weight on undercutting considerations – banks set "high" rates with lower probability. The higher bank lending ceilings, the lower then the average lending rate.

In the absence of CRT opportunities, bank's lending restraint is replaced by $\frac{K_i}{c}$. For any given R_i, $\mathcal{L}_i^{RT} > \mathcal{L}_i^D$, and $\mathcal{L}_i^D \geq \frac{K_i}{c}$ (by (22)), therefore credit rationing is less likely and the first-best optimum is more likely when CRT opportunities are available. By the same reasoning, in the mixed strategy regime, the expansion of banks' lending ceilings caused by CRT opportunities lowers the average lending rate.

Proposition 6 *CRT opportunities enhance bank competition for loans and always make borrowers better off. Banks are worse off, unless aggregate bank capital is sufficiently low so as to lead to credit rationing, i.e. condition* (23) *holds, in which case banks too are better off.*

If aggregate bank capital is sufficiently low so that inequality (23) holds, then in equilibrium i, $\forall i$, lends at the monopoly rate x and its lending volume equals the maximum it finds it optimal to monitor under optimal CRT. In the absence of CRT opportunities, $i's$ lending volume is further constrained: $i's$ profit per lending unit is the same as that attained when risk is transferred, but its lending volume is strictly lower (because, for any given R_i, $\frac{K_i}{c} \leq \mathcal{L}_i^D < \mathcal{L}_i^{RT}$). Thus, if the banking system is poorly capitalized, if condition (23) holds, then CRT opportunities not only benefit borrowers (firms) but also banks.

It is the *possibility* of transferring risk that makes competition tougher, as Proposition 6 correctly states. And, indeed in equilibrium not all banks will necessarily transfer risk. For instance, in the mixed strategy regime (i.e., when neither (23) nor (24) holds) or in the zero-profit regime (i.e. when (24) holds), a bank i whose lending volume turns out not to exceed $\frac{K_i}{c}$, may abstain from CRT. However, it is only if the banking system is sufficiently capitalized so as to lead to zero-profits in the absence of CRT, that no CRT takes place.[7]

Proposition 7 *CRT frees up capital: In an equilibrium where bank excess return on capital is strictly positive CRT necessarily occurs. Bank excess return on capital and CRT activity are positively correlated.*

Bank's excess return on capital is strictly positive whenever bank capital is not sufficiently abundant, i.e. condition (24) fails to hold. If aggregate bank capital is sufficiently scarce so as to lead to the monopoly regime, that is, if condition (23) holds, then each bank lends at the monopoly rate, is capacity constrained, i.e. $L_i = \mathcal{L}_i^{RT}$, $\forall i$, and transfers risk. Both CRT activity and bank excess return on capital attain their maximum levels. If aggregate bank capital is neither scarce nor abundant, that is if neither condition (23) nor (24) holds, then each bank randomizes its lending rate and earns a strictly positive excess return on capital, albeit lower than in the monopoly regime. Not all banks need to transfer risk, but necessarily there will be banks that transfer risk – certainly, the (low-rate setting) banks that despite CRT are capacity constrained. That is, both CRT activity and bank excess return on capital are strictly positive, and are lower than in the monopoly regime. As aggregate bank capital falls and/or aggregate demand for lending expands, more banks end up being capacity constrained (in the monopoly regime all banks are constrained), both bank excess return on capital and CRT activity increase – i.e. they are positively correlated.

8 Concluding Remarks

The key point is that (optimal) risk management leads to better alignment of rewards and efforts: The bank is insured against exogenous risk (bad luck)

[7] The necessary condition for no-risk transfer is that condition (24) holds, once $\mathcal{L}_z^{RT}$, is replaced by $\frac{K_z}{c}$, $z = i, j$, $j \neq i$.

and is rewarded/punished for what it is responsible: Monitoring incentives are enhanced. The practical/relevant implication is that the amount of outside finance the bank can raise expands. That is, banks can lever up capital to a greater extent: for any given amount of capital, the lending volume bank finds it incentive-compatible to monitor expands. Or, equivalently, optimal risk management allows banks to economize on capital: any given amount of incentive-compatible lending can be sustained with a lower amount of capital.

Diversification is the risk-management tool for idiosyncratic risk, (optimal) credit risk transfer (CRT) is the risk management tool for common risk. Indeed, when project returns are i.i.d. across firms and there is no common (aggregate) risk, the return of a diversified portfolio conditional on bank action (more generally, the agent's action) is certain: Diversification is sufficient for a debt-financed bank to find it optimal to monitor, the first best solution is attained (Diamond (1984), Cerasi and Daltung (2000), and in an industrial organization context, Laux (2001)). Aggregate risk, i.e. the fact that (otherwise i.i.d.) project returns depend on common factors, implies that the portfolio return conditional upon bank action is uncertain, even for a fully diversified bank. That is, banks are necessarily exposed to exogenous risk, even if perfectly diversified. For such an economy, we have shown that: i) monitoring-incentive compatibility potentially constrains the bank to inject a minimum amount of capital per lending unit, and thereby entails a potential constraint on bank's amount of outside finance and lending; ii) an optimal contract for the bank to raise finance makes use of the information conveyed by loan portfolio outcome and rewards the bank as much as possible for the outcomes that signal monitoring: The amount of capital per lending unit the bank needs to inject to find it incentive-compatible to monitor attains the minimum, bank's incentive-based lending capacity attains the maximum level; iii) debt financing is sub-optimal. Monitoring is under-rewarded: Bank faces a tighter constraint on the amount of outside finance it can raise (i.e., is credit rationed). Incentive-based lending capacity is smaller. However: iv) there exists (optimal) CRT such that raising outside finance with deposits/debt and then transfer the (exogenous) risk of the loan portfolio, i.e. engage in optimal CRT, is by all means equivalent to raising finance with an optimal contract; v) optimal CRT's reference asset is loan portfolio. It makes use of the information conveyed by loan portfolio outcome so as to insure the bank against bad luck (bad realizations of the common, macroeconomic, risk factor), if and only if loans are monitored: The amount of capital per lending unit the bank needs to inject to find it incentive-compatible to monitor attains the minimum, bank's incentive-based lending capacity attains the maximum level; vi) optimal CRT can be implemented by loan-portfolio securitization. Security holders are given (off-equilibrium threat) put options: The bank retains endogenous risk, that controlled by monitoring, and is insulated from the exogenous one. This would not be true for a partial loan sale (as in Gorton and Pennacchi, 1995; Pennacchi, 1998). In a partial loan sale, or equivalently a loan sale with partial recourse, the bank is necessarily exposed to exogenous risk – incentive-based capacity is smaller.

For simplicity, we have assumed that a bank's portfolio is sufficiently diversified to lead to a zero variance of the portfolio return conditional on the common factor (macro-state) realization, and on the bank action. In such a set-up, an optimal CRT scheme, loan securitization cum put options, insulates the bank from exogenous risk if and only if it monitors. With an undiversified portfolio, a monitoring (non-monitoring) bank may still be unlucky (lucky), i.e. the return realization of a monitored (unmonitored) portfolio may turn out to be sufficiently low (high) that options are (not) exercised: an undiversified bank is necessarily exposed to exogenous risk. To find it incentive-compatible to monitor, it needs to pledge an amount of capital per lending unit higher than that required for a diversified bank. Thus, also with aggregate risk, diversification is an important risk-management tool: the greater bank's diversification, the more it can be insulated from exogenous risk – the smaller the amount of capital it needs to pledge.

The model allows for one class of projects (firms) whose returns conditional on bank monitoring choice depend on a common factor, the macro-state realization (up-turn/down-turn). With more than one class, the equilibrium solution depends on the correlation of downside risks. Suppose that the economy consists of two sectors, 1 and 2. If the up-turn of sector 1 coincides with the down-turn of sector 2, i.e. downside risks are negatively correlated, then a (composite) portfolio of loans made to firms that belong to sector 1 *and* to sector 2 may be free of exogenous risk, in which case retaining all risks is an equilibrium whenever prudential (risk-based) regulation rewards diversification (specifically, the negative correlation of asset returns) – as Basel II does, by contrast to Basel I. If downside risks are positively correlated, i.e. when sector 1 is in the downturn sector 2 is in the downturn too, then a loan portfolio is necessarily exposed to exogenous risk. In an equilibrium, exogenous risk is transferred along the lines here derived: sector loans are pooled and securitized, and if sector loan portfolios are sufficiently diversified then banks will be insulated from exogenous risk. This also suggests that a bank industry structure where banks specialize in their lending is an optimal structure. A bank that specializes in the lending to a given sector holds a within-sector diversified portfolio whose exogenous risk can be (perfectly) disentangled and transferred. By contrast, bank organization design, for example a subsidiary structure that isolates loan portfolios cannot help, it cannot insure the bank against exogenous risk. Subsidiary structures help when asset returns are exogenous, not sensitive to monitoring/screening choices, in which case a subsidiary structure that isolates safer assets is more immune to risk-shifting than a unitary structure (Khan and Winton, 2004).

However, with uncoordinated depositors, for banks to engage in CRT, i.e. for value/welfare gains to obtain, a condition must hold: capital requirement on loans whose risks are retained (i.e. loans that are not securitized, or more generally, whose risks are not optimally transferred). Indeed, our further results are that, with dispersed depositors/bondholders, banks have incentive to over-loan capital, abstain from CRT, not monitor and shift losses on to lenders (depositors/bondholders). This bank moral-hazard problem is solved by prudential regulation: (incentive-based) capital requirements on loans whose risks are retained. If this condition is satisfied, then banks have incentives to engage in CRT – value/welfare gains obtain.

Finally, CRT opportunities have relevant effects on banks' strategic interactions in the loan market, and thereby on real investment activity. The possibility of transferring exogenous risk expands bank lending capacity: it makes loan competition fiercer, increases real investment, makes borrowers (firms) better off and banks worse off. Banks would benefit by committing not to engage in CRT, but this commitment is not credible, it violates individual rationality. Individual rationality calls for freeing up capital, which is what we find CRT allows for: If bank capital has economic value, i.e. if an equilibrium is one where bank excess return on capital is strictly positive, then necessarily banks transfer (exogenous) risk.

References

Arping, S. (2004), "Credit Protection and Lending Relationships", mimeo, University of Amsterdam.

Bank for International Settlements (2003), "Credit Risk Transfer", Committee on the Global Financial System, Basel.

Bank for International Settlements (2005a), "Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in 2004", Basel.

Bank for International Settlements (2005b), "Credit Risk Transfer", Basel Committee on Banking Supervision - The Joint Forum, Basel.

Bhattacharya, S., and G. Chiesa (1995), " Proprietary Information, Financial Intermediation and Research Incentives", *Journal of Financial Intermediation* 4, 328-357.

Bhattacharya, S. and A. Thakor (1993), "Contemporary Banking Theory", *Journal of Financial Intermediation* 3, 2-50.

Behr, P. and S. Lee (2005), "Credit Risk Transfer, Real Sector Productivity, and Financial Deepening", mimeo, Goethe University Frankfurt.

British Bankers Association (2004), "Credit Derivatives Report" 2003/04", London.

Campbell, T.S. and W.A. Kracaw (1980), "Information Production, Market Signalling and the Theory of Financial Intermediation", *Journal of Finance* 35, 863-882.

Carlstrom, C. T. and K.A. Samolyk (1995), "Loan Sales as a Response to Market-based Capital Constraints", *Journal of Banking and Finance* 19, 627-646.

Cebenoyan, S., and P.E. Strahan (2004), "Risk Management, Capital Structure and Lending at Banks", *Journal of Banking and Finance* 28, 19-43.

Cerasi, V., and S. Daltung (2000), "The Optimal Size of a Bank: Costs and Benefits of Diversification", *European Economic Review* 44, 1701-1726.

Chiesa, G. (1998), "Information Production, Banking Industry Structure and Credit Allocation", *Research in Economics* 52, 409-430.

Chiesa, G. (2001), "Incentive-Based Lending Capacity, Competition and Regulation in Banking", *Journal of Financial Intermediation* 10, pp.28-53.

Datta, S., M. Iskandar-Datta, and A. Patel (1999), “Bank Monitoring and the Pricing of Corporate Debt”, *Journal of Financial Economics* 51, 435-449.

De Marzo, P. (2005), "The Pooling and Tranching of Securities: A Model of Informed Intermediation", *Review of Financial Studies* 18, 1-36.

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

Duffee, G.R. and C. Zhou (2001), "Credit Derivatives in Banking: Useful Tools for Managing Risk?", *Journal of Monetary Economics* 48, 25-54.

Fama, E.F. (1985), "What's Different about Banks?", *Journal of Monetary Economics* 15, 29-39.

FitchRatings (2004), "Global Credit Derivatives Survey", FitchRatings Special Report, September 2004.

Froot, K.A. and J.C. Stein (1993), "Risk Management: Coordinating Corporate Investment and Financing Policies", *Journal of Finance* 48, 1629-1658.

Froot, K.A. and J.C. Stein (1998), "Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach", *Journal of Financial Economics* 47, 55-82.

Gorton, G.B. and G.G. Pennacchi (1995), "Banks and Loan Sales Marketing Nonmarkeatable Assets", *Journal of Monetary Economics* 35, 389-411.

Greenbaum, S.I. and A.V. Thakor (1987), "Bank Funding Modes: Securitization versus Deposits", *Journal of Banking and Finance* 11, 379-401.

Hellwig, M. (1991), "Banking, Financial Intermediation and Corporate Finance, in, European Financial Integration (Giovannini A. and C. Mayer, Eds.), pp. 35-63. Cambridge Univ.Press, Cambridge.

Holmstrom, B. and J. Tirole (1997), "Financial Intermediation, Loanable Funds and the Real Sector", *Quarterly Journal of Economics* 112, 663-91.

International Monetary Fund (2002), "Global Financial Stability Report", Washington.

James, C. (1987), "Some Evidence on the Uniqueness of Bank Loans", *Journal of Financial Economics* 19, 217-235.

Kahn, C. and A.Winton (2004), "Moral Hazard and Optimal Subsidiary Structure for Financial Institutions", *Journal of Finance* 59, 2531-2575.

Keenan, S. (2000), "Historical Default Rates of Corporate Bond Issuers 1920-1990", Working Paper, Moody's Investors Services.

Laux, C. (2001), "Limited-liability and Incentive Contracting with Multiple Projects", *RAND Journal of Economics* 32, 514-526.

Lummer, S.L., and J.C. McConnel (1989), “Further Evidence on the Bank Lending Process and the Capital Market Response to Bank Loan Agreements”, *Journal of Financial Economics* 25, 99-122.

Morrison, A.D. (2005), "Credit Derivatives, Disintermediation and Investment Decisions", *Journal of Business* 78, 621-647.

Nicolo’, A. and L.Pelizzon (2005), "Credit Derivatives: Capital Requirements and Strategic Contracting", mimeo, University of Venice.

Parlour, C. and G.Plantin (2005), "Credit Risk Transfer", mimeo, Tepper School of Business, Carnegie Mellon University.

Pennacchi, G. G.(1988), "Loan Sales and the Cost of Bank Capital", *Journal of Finance* 43, 375-395.

Ramakrishnan, R.T.S. and A.V. Thakor (1984), "Information Reliability and a Theory of Financial Intermediation", *Review of Economic Studies* 51, 415-432.

Wagner, W. and I.W.Marsh (2004), "Credit Risk Transfer and Financial Sector Performance", CEPR Discussion Paper No.4265.

Appendix A: Proof of Lemma 2

Let bank's lending be L, financed partly with capital and the remaining, $L-K$, with debt raised at a gross rate equal to one (which is always true in equilibrium, by the same reasoning as in Proposition 2). Let the bank securitize the loan portfolio, it sells L securities, each pays $\frac{1}{L}$ of the loan-portfolio return at the final date, carries the put option at (ii), and is issued at the price $S^m \equiv sR$. Bank's undominated strategy is either to monitor all loans, $a^* = m$, or no loan, $a^* = \varnothing$ (by the same reasoning as in Section 4). If $a^* = m$, then: i) in state $\underline{\theta}$ the security market price is αR, and options are not exercised; ii) the bank is necessarily solvent in $\underline{\theta}$ (by the same reasoning as in Section 4). Therefore:

$$\pi\left(a^* = m\right) = p\left[S^m L - (L-K)\right] + (1-p)\left[S^m L - (L-K)\right] - FL - K \ . \tag{B.1}$$

If $a^* = \varnothing$, then: i) in state $\underline{\theta}$ the security market price is $\underline{\alpha}R$, options are exercised, and bank's obligations to security holders amount to $RL\left(\alpha - \underline{\alpha}\right)$; ii) (non-monitoring) bank is necessarily insolvent in $\underline{\theta}$ (again, by the same reasoning as in Section 4). Hence:

$$\pi\left(a^* = \varnothing\right) = p\left[S^m L - (L-K)\right] + (1-p)\Pi\left(\underline{\theta}|a^* = \varnothing\right) - K \ , \tag{B.2}$$

where:

$$\pi\left(\underline{\theta}|a^* = \varnothing\right) \equiv \max\left[S^m L - (L-K) - RL_i\left(\alpha - \underline{\alpha}\right), 0\right] \equiv 0 \ .$$

The bank monitors if and only if $\pi\left(a^* = m\right) \geq \pi\left(a^* = \varnothing\right)$; that is, by using $(B.1) - (B.2)$ and $S^m \equiv sR$, if and only if:

$$\frac{K}{L} \geq -\left(sR - 1\right) + \frac{F}{1-p} \ ,$$

or equivalently, $L \leq \mathcal{L}^{RT}$. This establishes that condition (ii) is sufficient, and condition (i) is necessary and sufficient. That (ii) is also necessary follows because if (ii) fails to hold, i.e. securities do not carry put options, then $\pi\left(a^* = \varnothing\right) = \pi\left(a^* = m\right) + FL$, which clearly implies that the bank does not monitor.

We have then proved that if conditions (i)-(ii) hold, then $a^* = m$ and the security price $S^m \equiv sR$ provides final investors with zero profits, and if either (i) and/or (ii) fail to hold, then $a^* = \varnothing$ and the security price such that final investors make zero profits is $S^{\varnothing} \equiv \underline{s}R$. Lemma 2 is true ■

Table 1
Project Return Distribution

action\ *state*	$\overline{\theta}$	$\underline{\theta}$
m	$\Pr(X = x) = 1$	$\Pr(X = x) = \alpha$
$\varnothing$	$\Pr(X = x) = 1$	$\Pr(X = x) = \underline{\alpha}$