Bank incentives and optimal CDOs

Henri Pagès*
Banque de France, DGEI

July 29, 2009

Abstract

The paper examines a delegated monitoring problem between investors and a bank holding a portfolio of correlated loans displaying "contagion." Moral hazard prevents the bank from monitoring continuously unless it is compensated with the right incentive-compatible contract. The asset pool is liquidated when losses exceed a state-contingent cut-off rule. The bank bears a relatively high share of the risk initially, as it should have high-powered incentives to monitor, but its long term financial stake tapers off as losses unfold. Liquidity regulation based on securitization can replicate the optimal contract. The sponsor provides an internal credit enhancement out of the proceeds of the sale and extends protection in the form of weighted tranches of collateralized debt obligations. In compensation the trust pays servicing and rent-preserving fees if a long enough period elapses with no losses occurring. Rather than being detrimental, well-designed securitization seems an effective means of implementing the second best.

Keywords: Credit risk transfer, Default Risk, Contagion

JEL classification: G21, G28, G32

*Banque de France, DGEI, 49-1431, *E-mail address:* henri.pages@banque-france.fr. I would like to thank Viral Acharya, Gilles Chemla as well as participants in an internal seminar and the Finance Seminar at IDEI for helpful comments. I am especially indebted to Bruno Biais and Jean-Charles Rochet for pointing out a flaw in the previous version and making invaluable suggestions for improvement. All errors are mine. The usual disclaimers apply.

1 Introduction

Banks develop specialized monitoring skills on behalf of investors in exchange for investors' ability to fund their lending activies. Failure to commit adequate monitoring results in low credit standards, which can ultimately jeopardize financial stability. This buy-side agency problem can be felt acutely if loan losses are the result of contagious defaults when fundamentals go wrong. The focus of the paper is on how the contractual arrangements between banks and investors interact with banks' strategic behavior in determining credit standards and long term risk sharing when contagion is present.

One novel feature of the approach is to show that the delegated monitoring problem between a bank and risk-neutral investors can best be viewed in the context of asset-backed securities (ABS). In principle, the definition of ABS refers to a discrete pool of assets that self-liquidate under the passive purview of the trust in whose name the ABS are issued. If such were the case, static information about the timing and amount of expected payments would be enough to determine the performance of the pool, and there would barely be any servicing or management function to describe. However, the complex nature of ABS transactions introduces a lot of flexibility to administer the pool. One reason is that the pool may contain up to 50% of delinquent assets and compliance with the servicing agreement for the transaction is critical. Another is that active management of the pool is possible through the use of master trusts, prefunding periods or revolving periods, so that asset substitution becomes possible within certain limits.

While regulatory authorities have adopted specifically designed disclosure requirements to meet investors' concerns and foster transparency in ABS markets, the scope for moral hazard on the part of the servicer can be as important to the performance of the pool as its initial composition and characteristics. The second-best arrangement arrived at in the paper is consistent with the increasing realization laid out in the Federal Register (2005) and other references given herein (e.g., Section IIIB) that the servicing role in ABS transactions materially impacts the performance of the pool. We call such a role over the life of a porfolio of loans — including collection and management functions — "continuous monitoring."

To shed light on the dynamic delegated monitoring problem, we start with a stylized model where the bank may engage in unobservable actions that result in private benefits at the expense of performance.

We abstract from imperfect commitment problems and focus on moral hazard in risk prevention. More specifically, the bank can make a costly effort at any point in time to handle delinquent loans or manage its asset pool efficiently, in which case the portfolio's default intensity improves at that time. Given competitive investors, the goal is to elicit which high-powered compensation maximizes the bank's payoff subject to a zero-profit condition for investors and an incentive compatibility condition for continuous monitoring. The optimal trade-off between efficient risk sharing and efficient monitoring allows the bank to release as much costly capital as possible by laying off some credit risk while maintaining contract enforcement over the loans throughout.

Although the paper considers a single bank, systemic risk is handled with a model of "contagion." The model is Markovian, with individual default intensities depending on the number of non-performing loans. Consistently with the empirical evidence documented by Laurent at al. (2007), contagious dependence between defaults is introduced by assuming that the smaller the size of the portfolio, the higher individual risk. Correlation between default times comes from the fact that individual risk is *not* idiosyncratic. Each defaulting loan creates an externality on market participants' views about the quality of the rest of the portfolio. When contagion has spread, individual risk is extreme and it seems like losses are lumped together.

To simplify the exposition we consider a static portfolio of identical long term loans yielding constant cash flows per unit time. The optimal risk prevention policy relies on two instruments: positive payments to the bank and the threat of stochastic liquidation. In line with the growing literature on dynamic moral hazard, these decisions are made on the basis of two state variables: the size of the portfolio and the continuation utility of the bank. While the former reflects the total number of losses, the latter summarizes the track record of performance. The two must be distinguished because the assessment of performance relies on how quickly the portfolio has unraveled, not how much. We characterize the compensation and stochastic liquidation policy arising from the optimal contract.

Consider first the compensation policy. In order to have the bank work in their best interest, investors resort to the carrot-and-stick approach. The bank is rewarded when its track record is on target. Two kind of fees are charged in the "bliss" state. One is the *servicing* fee for monitoring which is a flat percentage of the portfolio size. The other is the *rent-preserving* fee for impatience which depends on the bank's discount rate. When the track record deteriorates, however, payments are suspended. The bank takes stick from investors through a reduction in its continuation utility as soon as a loss occurs. The magnitude of the "punishment" is pinned down by the incentive compatibility constraint. In the beginning underlying risk is low and it is difficult to disentangle a bank which monitors from one which does not. It needs high-powered incentives and bears the brunt of initial losses. In the end underlying risk is high and the imminence of a default makes the bank eager to monitor. It is no longer tantalized by the prospect of shirking and better shielded against the incidence of losses in financial distress. Thus, according to this compulsory retention scheme, the bank's *risk share* tapers off as losses unfold, until the portfolio is exhausted.

Next consider the stochastic liquidation policy. Punishments meted out during compensation define the bank's reservation utility. If the target set in the bliss state comes close to reservation utility, the threat of reductions has no real bite because the bank is protected by limited liability. To cope with the situation, investors allow for random liquidation[1] of the portfolio upon default, with a probability of survival reflecting the bank's current performance. The threat of liquidation impels the bank to keep monitoring when its performance is poor but is socially costly, so investors are keen to keep stochastic liquidation as far as possible from target. The gap between the best and worst performances for given size defines a contingent *cut-off rule*. It is the highest permissible level for losses starting from bliss or, more precisely, the maximum number of joint defaults that the bank is allowed to make without fearing liquidation. Tuning the cut-off rule is as effective an instrument to discipline banks as the punisment itself.

To understand the mechanics of the cut-off rule, recall that once on probation the bank is driven by the prospect of future payments. As long as there are no losses, payments should be resumed soon and the new target adjusted to assuage the bank's desire for fees. In normal circumstances — assuming individual risk is not noticeably affected — it is not sensible to keep the bank waiting with the promise of larger payments since underlying conditions have not changed. The reason for actually reducing payments is twofold. First,

[1] An alternative threat against a non-performing bank would be downsizing the portfolio. Although this would achieve essentially the same outcome, the implementation might be more difficult if loans are indistinguishable. Recurrent downsizing could also be viewed as more disruptive.

the bank's compensation should not improve in size-adjusted terms and the portfolio has decayed by one unit. Second, the bank's risk shifting incentives should be held in check and losses are slightly less frequent, making shirking more difficult to detect. On both accounts, investors' best reaction is to lower the target by strengthening the cut-off rule. Thus, looking forward from the preceding state, the well-performing bank knows that it will lose after a loss even if it does not have to wait long and remains diligent.

But there is a twist. Individual risk may surge in rare circumstances. It is then that the bank's special skills at collecting loan payments are most valuable. The aggregate loss intensity soars despite the reduction in portfolio size and dwarfs the bank's own discount rate, making the cost of performance appear relatively cheap. Investors' best reaction is to rescale the number of permissible losses to take the new conditions into account, i.e., slacken the cut-off rule. By this token, heightened concerns about underlying risk induce an abrupt fall in reservation utility, but their impact on the target is dampened. Again looking forward from the preceding state, the well-performing bank knows that it will have to wait a long time before payments are resumed if it comes to operate under turbulence and so remains diligent.

Interestingly, the optimal prevention policy can be implemented through a true sale transaction when changes in underlying risk are lumpy. To this extent the complex institutions involved in structured finance can sow the seeds of their salvation. More specifically, the sponsoring bank sells its portfolio to a trust and guarantees the deal by returning the capital required and gain on sale to a reserve account managed by the trust. The internal credit enhancement is used as cash collateral to reimburse the trust for losses following tranches of protection matching the bank's optimal risk shares. Fees accrue on the reserve account to increase credit support and are remitted when the balance is on target. In contrast, premium spreads are retained by the trust as a liquidity tax prepaid by the sponsoring bank for the systemic risk it creates. The idea is that movements in the reserve account balance faithfully mirrors the bank's performance and can be used by the trust to trigger stochastic liquidation. The result shows that, with continuous monitoring, the optimal tradeoff between risk sharing and monitoring is consistent with separating different functions in the production process, with origination and servicing on the one hand, and activities related to securitization on the other.

The paper belongs to the recent and fertile literature on dynamic moral hazard, as illustrated by DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007a, 2007b), Biais et al. (2007) or Sannikov (2008). Many papers deal with frequent and infinitesimal risk, but Sannikov (2005) also has Poisson risk. A difference is that jumps are associated with upside cash-flow shocks, which leads to predictable downsizing and qualitatively different results. In Biais et al. (2009), moral hazard is about large and infrequent risks. As in our model and unlike in the Brownian case, investors inflict sharp reductions in the agent's continuation utility when losses occur and unpredictable downsizing when performance is poor. Firm size dynamics is markedly different because the agent can expand through investment and follow asymptotically a positive growth trend. In contrast ABS refer to a discrete pool of assets that eventually ceases to exist. Our analysis offers a first description of unpredictable downsizing in a non-stationary context.

The paper is related to several other strands of literature. One deals with the importance of forward monitoring in banking using continuous flow of information. Peeking at the checking account balance or financial statements helps banks monitor outstanding loans as outlined in Norden and Weber (2008). Dichev and Skinner (2001) argue that banks set loan covenants very tight and use them to work with borrowers behind on payments, possibly extending grace periods and paring fees or interest rates to minimize losses. There is also evidence about the importance of servicers in securitization. Ashcraft and Shuermann (2008) show that the servicer's role is not confined to the collection and remittance of loan payments and carries important responsibilities, like maintenance of property, hazard insurance and tax bills when the loan starts being delinquent or like prompt foreclosure once deemed uncollectible. These activities have consequences for the distribution of cash flows, with an impact of plus or minus 10 percent on losses according to one Moody's estimate. Gan and Mayer (2006) discuss the role of the "special servicer" who is responsible for the borrower work-out and foreclosure functions. They find that when they hold the first-loss piece, special servicers appear to behave more efficiently, with a positive impact on the price of junior tranches. In Cantor and Hu (2006), the weaker performance of certain types of sponsors is related to their incentives to economize on quality servicing or select risky assets.

Several papers examine the implications of credit risk transfer (CRT) for banks' incentives to monitor that recent empirical studies document.[2] They generally find that CRT has negative repercussions on monitoring incentives. These results hold against the backdrop of Innes (1990), who shows that under a monotone likelihood ratio property (MLRP) debt financing maximizes the reward for monitoring. A notable exception is Chiesa (2008), who departs from MLRP by assuming that the medium performance of a portfolio must reveal a bank that has monitored in a downturn. In her paper, good performance is always attributed to good luck and monitoring is only useful in downturns. Fender & Mitchell (2008) extend the model in various dimensions to focus on the incidence of different retention mechanisms on banks' incentives to screen borrowers. Here we suggest that the lack of MLRP is not necessary to vindicate CRT. In Arping (2004), credit derivatives have a positive impact on incentives because they insulate banks from default risk before maturity and promote early efficient liquidation, which strengthens borrowers' incentives for effort. A few papers study how different forms of CRT affect the efficiency of monitoring. In Duffee and Zhou (2001), introducing credit derivatives promotes risk sharing but undermines the loan-sale market. The effect on monitoring depends on whether or not there is pooling in the loan-sale market. Parlour and Winton (2008) consider the value of control rights in the loan-sale market when loan sales and credit derivatives coexist. They find that none of the equilibria can achieve both efficient risk sharing and efficient monitoring. These papers do not consider partial credit enhancements associated with the provision of continuous monitoring.

Pooling and tranching have been rationalized in the literature, in particular as an incentive for issuers to acquire inside information about asset values prior to sale. Using the security design model of De Marzo and Duffie (1999), DeMarzo (2005) shows that tranching mitigates an adverse selection problem by allocating information-insensitive derivative securities to uninformed investors while intermediaries' retention of junior tranches signals their superior ability in valuing assets. In a similar vein, Plantin (2004) shows that tranching is optimal when financial institutions differ in their ability to screen collateral and redistribute securities. A paper close to ours is Franke and Krahnen (2006) who argue that, with payoffs indexed to system-wide macroeconomic shocks, senior tranches are better held by investors with no relationship-specific information, while intermediaries' retention of junior tranches ensures that their risk share increases with the influence

[2] See, for example, Berndt and Gupta (2008), Drucker and Puri (2007), Keys et al. (2008).

they have through monitoring. Interestingly, their results indicate that banks' securitization activity is associated with an increase in their systematic risk, not a reduction, which they interpret as the higher correlation in risk exposures implied by banks reinvesting the proceeds from securitization in new loans with the same properties as those in their initial books. Duffie (2008) uses numerical simulations to show that the issuer has an incentive to reduce dramatically both the fraction retained and the effort level when the cost of monitoring is sufficiently high. The reduction in default intensity through monitoring follows Duffie and Gârleanu (2001) and features a richer set of parameters and controls than in our model. On the other hand, retention is limited to the equity tranche.

The paper is organized as follows. In Section 2, we present the model and characterize the optimal contract under the incentive compatibility and limited liability constraints. Based on this analysis, Section 3 adopts a backward recursive approach to construct the solution of a system of optimal control problems and derive the dynamics of bank size. Section 4 offers some tentative policy implications before the conclusion in Section 5. Proofs are simple and left for the reader, except for that of Proposition 3 provided in Appendix.

2 The Model

The economy consists of a bank and many investors. The former has sole rights to the returns of relationship-specific loans. It has limited liability, some funds to start with, and derives private benefits from not conducting due diligence on the loans. The latter have unlimited liability and supply liquidity competitively, as long as they cover the costs. There is universal risk neutrality and the bank is more impatient than investors. Investors' objective is to find a contract that maximizes their expected profit subject to the bank maintaining loan enforcement and getting its reservation utility. Knowing this, the bank will offer investors a contract that allows it to fetch as much as it can and let them break even.

The bank's opportunity set consists of I unit loans, the default risk of which has some systematic component. All loans are ex ante identical and yield μ per unit time. The portfolio is static, with no reinvestment after time zero. However, when a loan gets repaid, it is immediately replaced by a loan with the same characteristics. Investors can commit to liquidate the pool in case of poor performance, but loans are worth nothing if managed outside the bank. This is meant to capture the idea that a bank's portfolio illiquidity stems in large part from the bank-customer relationship, implying that the ability to collect loans rests squarely on the lender's unique skills at working with borrowers behind on payments or extracting more concessions from them.

Monitoring is often viewed as the choice of costly effort made by a lender at origination to screen borrowers in an adverse selection environment. In this paper, we emphasize instead the choice of costly effort dedicated by one or more *servicers* during the life of the loans to support a deteriorating performance. For example, a bank can set debt covenants whose fulfillment is then monitored. The Federal Register (2005) shows that servicing is often quite complex in securitization and can entail a division of responsibilities between several entities: a "master servicer" oversees the action of other servicers, "primary servicers" are responsible for primary contact with obligors and collection efforts, "special servicers" are charged with handling borrower work-out and foreclosure functions, while an "administrator" is entrusted with the dynamic management, possibly adding new units to the pool from funds set aside or recycled cash flows.

Such dynamic monitoring has two consequences. First, the distinction between the exogenous base quality of the loan and the endogenous default probability that obtains after the monitoring decision has taken place arises at each point in time. Second, the cost of monitoring depends on how defaults propagate in the portfolio. We rely on a homogeneous "contagion" Markov model where the loss intensity of the nth-to-default loan depends on the size of the pool. We show that, if investors can commit to liquidate loans before maturity, they will ensure that the bank is diligent by winding down the pool when losses exceed a state-dependent threshold.

2.1 Dynamic monitoring

Let $i = I - N_t$ be the size of the portfolio[3] at time t, where $N_t = 0, \ldots, I$ is the default count. Downsizing occurs either as a result of individual defaults or of liquidation by investors. The information $\mathcal{F}_t$ is the natural filtration associated with the default and liquidation times.

[3] To avoid cumbersome notation, the time index of portfolio size i is systematically suppressed.

The default count N_t is a controlled time-homogeneous and Markovian process (Karlin and Taylor, 1975). Under the risk neutral probability, the individual default indicators N_t^j have default intensities depending on the size of the pool and on the level of bank monitoring. If the bank monitors continuously, default intensities are $\alpha^j(t) = \alpha_i$ for the i loans outstanding and zero otherwise. Thus, as long as the bank is diligent and spared from liquidation risk, the aggregate loss intensity of the pool is $\lambda_i = \sum_j \alpha^j(t) = i\alpha_i$.

Monitoring effort is costly and unobservable to investors. It affects risk only at the time it is exerted. As in Holmström and Tirole (1997) there are two levels of effort. If the bank chooses to shirk ($e_t = 1$), it enjoys a private benefit $B\,dt$ per loan between t and $t + dt$, in which case the aggregate loss intensity, $(1 + \epsilon)\lambda_i$, is higher than what it would be under monitoring ($e_t = 0$), uniformy in i.

A contract specifies the amount δ_t to be paid to the bank and the time τ at which liquidation occurs, if ever. Liquidation is unpredictable and stochastic, as it takes place only after a loss and depends on the realization of a lottery. The survival probability given default is denoted by θ, so the pre-liquidation intensity associated with the indicator $M_t = 1_{\{t \geq \tau\}}$ is $\lambda_i(1+\epsilon e_t)(1-\theta)$. The sequence of events is as follows. The size inherited from the past is i. The bank receives payment $\delta_t\,dt$ and makes its effort decision e_t for $(t, t + dt)$. With probability $\lambda_i(1 + \epsilon e_t)\,dt$ there is a loss and the size becomes $i - 1$. Then the pool is liquidated with probability $1 - \theta$. Otherwise the bank keeps administering the pool, with initial size or one less unit.

2.2 Incentive compatibility and limited liability

Let r be the bank's rate of impatience. The interest rate, including any premium that investors pay for consuming early, is normalized to zero. As in Sannikov (2008) or Biais et al. (2009), we specify the bank's lifetime utility at t as the conditional expected discounted revenue of its activities

$$U_t = E\left[\int_0^\tau e^{-rs}\left(\delta_s + 1_{\{e_s=1\}}B(I - N_s)\right)\,ds \,\middle|\, \mathcal{F}_t\right],$$

given a contract (δ, τ) and an effort process e. Related to lifetime utility is the bank's continuation utility defined as

$$u_t = 1_{\{t \leq \tau\}} E \left[\int_t^\tau e^{-r(s-t)} \left(\delta_s + 1_{\{e_s = 1\}} B(I - N_s) \right) ds \,\middle|\, \mathcal{F}_t \right]. \tag{1}$$

The bank participates only if its continuation gains, plus any monetary and private dividends, match its impatience. Since

$$U_t = \int_0^{t \wedge \tau} e^{-rs} \left(\delta_s + 1_{\{e_s = 1\}} B(I - N_s) \right) ds + e^{-rt} u_t$$

is a martingale, the integral representation theorem for point processes (Brémaud, 1981) implies that there are predictable processes[4] h^1 and h^2 such that the bank's continuation utility satisfies the promise-keeping equation

$$du_t + (\delta_t + e_t B(I - N_t))\, dt = r u_t\, dt - h_t^1 \left(\sum_j dN_t^j - \alpha_t^j (1 + e_t \epsilon)\, dt \right) - h_t^2 \left(dM_t - (1-\theta)\lambda_i (1 + e_t \epsilon)\, dt \right) \tag{2}$$

until liquidation. The bank's expected change in continuation utility, net of payments and private benefits, is equal to r, while h^1 and h^2 are the sensitivities of utility to individual losses and liquidation, respectively. We have the following result, in line with Sannikov (2008, Proposition 2).

Proposition 1 *Given a contract (δ, τ), the effort process is incentive compatible if and only if*

$$h_t^1 + (1-\theta) h_t^2 \geq b_i = \frac{B}{\epsilon \alpha_i}, \tag{3}$$

almost surely for all $t \in [0, \tau]$.

Heuristically, if the bank plans to follow the optimal strategy $e = 0$ starting from t, it should have no incentive to deviate before t. From (1), its continuation utility u_t is determined by the history of defaults and the contract (δ, τ) after time t, not by its effort before time t. Given u_t, it will not deviate between $t - dt$ and t if the real change in continuous utility $du_t - r u_t\, dt$ is *lower* under monitoring. This yields the incentive

[4] Since oustanding loans are indistinguishable, we assume w.l.o.g. that h^1 is a scalar process.

compatibility constraint (3). The left-hand side is the predictable loss in utility brought about by default and liquidation risk. The right-hand side is the minimum rent consistent with monitoring under limited liability. Indeed, the promise-keeping and incentive compatibility conditions (2) and (3) taken together[5] imply that losses inflicted upon default can always be as high as b_i. Continuation utility cannot fall below b_i.

A high sensitivity to losses requires that the bank be compensated with high utility in the begining. This reduces investors' value. Hence, the incentive compatibility binds under the optimal plan. Because liquidation is inefficient and should be avoided to the extent possible, there are two regimes for the bank. Either $u \geq b_i + b_{i-1}$ and there is no need to liquidate the pool ($\theta = 1$). The loss in utility is $h_t^1 = b_i$ and since $u - b_i \geq b_{i-1}$ the limited liability constraint is not violated in state $i-1$. Or $b_i \leq u < b_i + b_{i-1}$ and liquidation is necessary. Since all is lost when the pool is wasted, the promise-keeping constraint yields $u = h^1 + h^2$. The incentive compatibility constraint in turn determines $\theta = (u - b_i) / (u - h_1)$. But limited liability has $u - h^1 \geq b_{i-1}$ when the pool is spared, so θ is maximized when $h^1 = u - b_{i-1}$ and $h^2 = b_{i-1}$. The optimal survival probability, $\theta = (u - b_i) / b_{i-1}$, reflects the bank's position in the interval $[b_i, b_i + b_{i-1}]$. If a default occurs, utility is first reduced to $u - h^1 = b_{i-1}$, the bank's reservation utility in state $i-1$. Then a coin is thrown. Heads the bank remains in charge and its utility starts growing. Tails the bank is dispossessed and $b_{i-1} - h^2 = 0$.

2.3 Optimal contracting

If $h^1 = b_i \wedge (u - b_{i-1})$, $h^2 = b_{i-1}$ and $\theta = (u - b_i) / b_{i-1} \wedge 1$, the contract is incentive compatible. The promise-keeping equation (2) returns

$$\begin{aligned} \dot{u} + \delta_t &= ru + \lambda_i b_i \wedge (u - b_{i-1}) + \lambda_i (1 - \theta) b_{i-1} \\ &= ru + \lambda_i b_i \end{aligned}$$

[5] If $\theta = 1$ there is no liquidation and $\Delta u = -h^1 = -h^1 - (1-\theta)h^2 \leq -b_i$. Otherwise $-\Delta u$ is either h^1 (probability θ) or $h^1 + h^2$ (probability $1 - \theta$). Since $\max\left(h^1, h^1 + h^2\right) \geq h^1 + (1-\theta)h^2 \geq b_i$, $\Delta u \leq -b_i$ with strictly positive probability.

between two successive losses. The bank charges two kinds of fees to investors. One shields the bank against the incidence of losses for which it is not accountable under monitoring. The servicing fee $\lambda_i b_i = iB/\epsilon$ is a flat percentage of the outstanding portfolio. The other maintains the real value of the bank's continuation utility. The rent-preserving fee ru is tuned to the rate of impatience.

In this time-homogeneous setup, as in many models of dynamic moral hazard, the current size of the portfolio i and the bank's current utility u are sufficient statistics for the optimal contract. Investors' continuation utility, $v_i(u)$, satisfies the following system of dynamic Hamilton Jacobi Bellman equations which can be solved recursively from $i = 1$ to I

$$\max_{\delta_t(\cdot)} \{\dot{v}_i\,(ru + \lambda_i b_i - \delta_t) + i\mu - \delta_t - \lambda_i\theta\,(v_i(u) - v_{i-1}((u - b_i) \vee b_{i-1}) - \lambda_i(1-\theta)v_i(u)\} = 0,$$

where $\theta = (u - b_i)/b_{i-1} \wedge 1$ is the optimal probability of liquidation given default. The first term is the change in continuation value brought about by the drift in u. The second is the revenue from the loans net of payment to the bank. The last two correspond to the loss of utility incurred depending on whether the bank keeps operating or not, respectively. With the extrapolation $v_i(u) = u/b_i\, v(b_i)$ on $u \in [0, b_i]$ the HJB equations can be simplified as

$$\max_{\delta_t(\cdot)} \{\dot{v}_i\,(ru + \lambda_i b_i - \delta_t) + i\mu - \delta_t - \lambda_i\,(v_i(u) - \theta v_{i-1}(u - b_i))\} = 0.$$

Movements in u reflect the history of individual losses: u keeps increasing towards some target unless some unexpected default brings it down. The complementary slackness condition $\delta_i\,(\dot{v}_i + 1) = 0$ helps explain why. When u is above target, social surplus $u + v_i$ is maximized and $\dot{v}_i = -1$. Investors prevent u from rising above target by paying fees to the bank. Below target $\dot{v}_i > -1$ and investors are better off postponing payments until the target is reached. A string of unexpected losses can interrupt this process. If u falls below $b_i + b_{i-1}$ in state i, the bank fears liquidation risk after a loss.

3 Bank size dynamics

With constant returns to scale, the bank's reservation utility, $b_i = B/(\epsilon\alpha_i)$, does not change as long as α_i remains constant. In a stable environment with constant individual risk, the size-adjusted rent $B/(\epsilon\lambda_i)$ rises despite bad performance at the rate $1/i$ (which is increasing, since i is decremented by the default count). Because the aggregate default rate declines with the number of loans outstanding, it becomes increasingly difficult to disentangle a bank which monitors from one which does not. Its informational rent per unit loan edges up. In contrast, when bouts of contagion trigger a sharp rise in the underlying default risk, the bank's size-adjusted rent falls abruptly. It has less leeway to shirk. We are interested in the implications that such changes have for the design of the optimal contract.

Contagion between defaults is introduced by assuming that the sequence α_i is low in the beginning and eventually high, i.e., $\alpha_I \leq \alpha_{I-1} \leq \cdots \leq \alpha_1$. This imperfect correlation between default times undermines the bank's ability to diversify its credit risk and makes the last few loans comparable to "economic catastrophe bonds" (Coval et al., forthcoming), low in risk unconditionnally but likely to be wiped out if the risk materializes.

We make the following assumptions.

Assumption 1 *The sequence α_i is decreasing (chronologically increasing): $\alpha_i \leq \alpha_{i-1}$.*

Since $\lambda_{i-1}/\lambda_i \geq (i-1)/i$, aggregate default intensity mirrors jumps in underlying individual risk and cannot decrease by more than $1/i$ if the latter is constant. A special case obtains when individual risk does not vary with size (local independence). Aggregate default intensity being proportional to size may be high in the beginning if the spell is long.

Assumption 2 $\inf_{i\geq 2} i\alpha_i > r$.

Aggregate intensity is higher than the bank's rate of impatience starting from $i = I$ to $i = 2$.

Assumption 3 *There exists k such that $k \leq \alpha_i/\alpha_{i-1} < 1$ and $\overline{w}_1 = \mu\epsilon/B - (\lambda_1 + r)/\lambda_1 > 1/\ln(1 + k/2)$.*

This technical condition requires that the loan revenue be all the higher in proportion of rents, the higher the maximum increase in individual risk $\alpha_{i-1}/\alpha_i = 1/k$.

3.1 Single loan: Constant utility

Bank's continuation utility is constant and set at its minimum level $\overline{u}_1 = b_1$. This implies a continuous payment of $\delta_1 = b_1 (r + \lambda_1)$ until extinction at time τ. Since $E\left[\tau - t \mid \mathcal{F}_t\right] = 1/\lambda_1$, investors' utility is

$$\overline{v}_1 = E \int_t^\tau (\mu - \delta_1)\, ds = \frac{\mu - \delta_1}{\lambda_1} = b_1 \overline{w}_1$$

with $\overline{w}_1$ as in Assumption 3. Optimal policy is captured by the value function $v_1(u) = b_1 w_1(\theta)$ with $\theta = (u - b_1)/b_1$ and normalized continuation utility

$$w_1(\theta) = \overline{w}_1 - \theta,$$

When $u > b_1$, an immediate payment of $u - b_1$ is made to have the bank fall back on its reservation utility b_1. However, $u > b_1$ is never reached under the optimal plan.

3.2 Two loans: Stochastic liquidation

In the absence of payments until the target is reached, the bank's continuation utility grows as

$$\dot{u} = ru + \lambda_2 b_2 = ru + 2B/\epsilon, \qquad u \in [b_2, b_2 + b_1).$$

Investors' continuation utility $v_2(u)$ satisfies the following HJB equation

$$\dot{v}_2 (ru + \lambda_2 b_2) + 2\mu - \lambda_2 (v_2 - \theta \overline{v}_1) = 0$$

where $\theta = (u - b_2)/b_1$is the survival probability given default. Hence between b_1 and b_2 there is stochastic liquidation, the bank's utility is $u = b_2 + b_1\theta$ and $1 - \theta$ is the probability that the pool is dissipated if a default occurs.

With normalized $w_2^1(\theta) = b_1^{-1} v_2(u)$ we have[6]

$$\dot{w}_2^1\left(\frac{b_2}{b_1}(\lambda_2 + r) + r\theta\right) + \lambda_2 \frac{b_2}{b_1}\frac{\mu\epsilon}{B} - \lambda_2(w_2^1 - \theta\overline{w}_1) = 0,$$

the solution of which can be written as

$$w_2^1(\theta) = B_2^1 + A_2^1\theta - C_2^1 \frac{\Lambda_2^1}{\lambda_2}\left(\frac{\Lambda_2^1 + r\theta}{\Lambda_2^1}\right)^{\lambda_2/r}$$

where parameters (also used in sections below) are as follows

Variable	Λ_i^j	A_i^j	B_i^j	y_i
Definition	$\frac{b_{i-j+1}}{b_{i-j}}\left(\Lambda_i^{j-1} + r\right)$	$\frac{\lambda_i}{\lambda_i - r} A_{i-1}^{j-1}$	$\sum_{k=0}^{j-1} \frac{b_{i-k}}{b_{i-j}}\frac{\mu\epsilon}{B} + A_{i-k}^{j-k}\frac{\Lambda_{i-k}^{j-k}}{\lambda_{i-k}}$	$\left(\frac{\Lambda_i^1}{\Lambda_i^1 + r}\right)^{\lambda_i/r-1}$
Remarks	$\Lambda_i^0 = \lambda_i$	$A_i^0 = \frac{b_{i-1}}{b_i}\overline{w}_i$	$B_i^1 = \frac{b_i}{b_{i-1}}\frac{\mu\epsilon}{B} + A_i^1\frac{\Lambda_i^1}{\lambda_i}$	$\frac{\mu\epsilon}{B} > \frac{\lambda_i - r}{\lambda_i}\frac{y_i}{1-y_i}$

with $i \geq 2$, $j < i$ and the convention $b_0 = b_1$.

The free parameter C_2^1 is determined by the boundary condition $\dot{w}_2(1) = -1$, yielding $C_2^1 = \left(1 + A_2^1\right) y_2$. Indeed, it is neither optimal to prevent the bank's continuation utility from increasing before the survival probability θ is equal to one, nor to keep it increasing beyond $u = b_2 + b_1$. Hence the stochastic liquidation region is exactly $[b_2, b_2 + b_1)$. In the absence of default, the target $u = b_2 + b_1$ is reached, there is no more risk of stochastic liquidation and the bank receives a continuous payment of $\delta_2 = r(b_2 + b_1) + 2B/\epsilon$ until either of the two loans defaults. Its continuation utility then jumps to b_1.

Investor's continuation utility is concave on $[b_2, b_2 + b_1)$. This property, as in all solutions of higher order states, reflects the inefficiency arising from stochastic liquidation. The principal's value react strongly to performance when liquidation is likely and much less so if the track record is good. The function w_2^1 also starts increasing under Assumption 3. The technical condition implies that $A_2 = \lambda_2 / (\lambda_2 - r)\,\overline{w}_1 > y_2 / (1 - y_2)$.

[6] The notation $w_i^j(\theta)$ refers to a normalized solution with i loans over the jth interval $\left[b_i + \cdots + b_{i-(j-1)}, b_i + \cdots + b_{i-j}\right]$. It is defined as $w_i^j(\theta) = v_i(u)/b_{i-j}$ where θ is the position of u in that interval.

Hence $C_2^1 = \left(1 + A_2^1\right) y_2 < A_2^1$ and $\dot{w}_2^1(0) = A_2^1 - C_2^1$ is positive. Moreover

$$\overline{w}_2 = w_2^1(0) = \frac{b_2}{b_1}\frac{\mu\epsilon}{B} + \frac{\Lambda_2^l}{\lambda_2}\left(A_2^1 - C_2^1\right) > \frac{b_2}{b_1}\frac{\mu\epsilon}{B},$$

a result needed just below.

3.3 Three loans: Probation

In the stochastic liquidation interval $[b_3, b_3 + b_2]$, investors' continuation utility is solved as before as

$$w_3^1(\theta) = B_3^1 + A_3^1\theta - C_3^1\frac{\Lambda_3^1}{\lambda_3}\left(\frac{\Lambda_3^1 + r\theta}{\Lambda_3^1}\right)^{\lambda_3/r}.$$

for some C_3^1. Anticipating somewhat one can show again that investors' utility starts increasing. Before the target is reached one must have $\dot{w}_3^1(1) > -1$ or equivalently $C_3^1 < \left(1 + A_3^1\right) y_3$. But we have just seen above that $\left(\lambda_3 - r\right)/\lambda_3 A_3^1 = b_1/b_2\overline{w}_2 > \mu\epsilon/B$ and, following Assumption 3, $\mu\epsilon/B > (\lambda_3 - r)/\lambda_3 y_3/(1 - y_3)$. It follows that $\left(1 + A_3^1\right) y_3 < A_3^1$ and $\dot{w}_3^1(0) = A_3^1 - C_3^1 > 0$. Moreover

$$\overline{w}_3 = w_3^1(0) = \frac{b_3}{b_2}\frac{\mu\epsilon}{B} + \frac{\Lambda_3^1}{\lambda_3}\left(A_3^1 - C_3^1\right) > \frac{b_3}{b_2}\frac{\mu\epsilon}{B},$$

a property preserved by induction across all stochastic liquidation intervals of higher order states.

It is no longer optimal to prevent u from exceeding the stochastic liquidation interval. Beyond $b_3 + b_2$ the bank must be let out on probation for some time. The bank's position, $\theta = \left(u_3 - b_3 - b_2\right)/b_1$, is the probability of survival given default that the bank finds should it lose a loan. The solution is

$$\begin{aligned} w_3^2(\theta) &= B_3^2 + A_3^2\theta + \phi_3^2(\theta) - C_3^2\frac{\Lambda_3^2}{\lambda_3}\left(\frac{\Lambda_3^2 + r\theta}{\Lambda_3^2}\right)^{\lambda_3/r}, \\ \left(\Lambda_i^j + r\theta\right)\dot{\phi}_i^j(\theta) &= \lambda_i\left(\phi_i^j(\theta) - w_{i-1}^{j-1}(\theta) + B_{i-1}^{j-1} + A_{i-1}^{j-1}\theta\right), \qquad \phi_i^j(0) = 0. \end{aligned}$$

The pasting condition

$$\dot{w}_3^1(1) = \dot{w}_3^2(0) \Longleftrightarrow A_3^1 - C_3^1/y_3 = A_3^2 - C_3^2 + \dot{\phi}_3^2(0)$$

specifies C_3^1 as a function of C_3^2. The differential equations with lags defining the optimal plan have solutions that are continuously differentiable. Pasting derivatives ensures that levels adjust.

Let $\overline{\theta}_3$ be the upper boundary of probation. If $\overline{\theta}_3 \in (0, 1)$, it solves the system

$$\begin{aligned} \dot{w}_3^2(\theta) &= -1 \\ \ddot{w}_3^2(\theta) &= 0. \end{aligned}$$

The first condition states that it is no longer cheaper to compensate the bank using future rewards than an immediate transfer. The second condition is a "smooth pasting" condition ensuring that $\overline{\theta}_3$ is optimal. Indeed, if w_3^2 were strictly concave at $\overline{\theta}_3$, more surplus could be obtained by marginally raising the treshold beyond that level. Differentiating the ODE defining ϕ_3^2 to eliminate $\ddot{\phi}_3^2$, one finds after some substitutions that the critical level $\overline{\theta}_3$ satisfies

$$1 + \dot{w}_2^1(\theta) = \frac{r}{\lambda_3}$$

The size of probation is determined by investors' marginal utility in the stochastic liquidation interval one step ahead. By construction the slope of the objective function declines to -1 until $\theta = 1$ so $\overline{\theta}_3$ is certainly less than one. It turns out in this rather special case that the slope is positive at $\theta = 0$. A positive slope signals a severe hazard moral problem since a better performance improves both the utility of the principal and that of the agent. There is thus room for improvement and $\overline{\theta}_3$ is away from zero. All parameters are then recovered from the boundary condition $1 + \dot{w}_3^2(\overline{\theta}_3) = 0$.

The higher the aggregate intensity in state 3, the larger probation. The intuition is the following. Aggregate default intensity is highest when individual risk remains constant ($\alpha_3 = \alpha_2$). But it improves with the downsizing, making opportunistic behavior more difficult to detect and raising the bank's size-adjusted rent in state 2. To restore the bank's incentives to monitor, investors must raise the stakes by lengthening probation and promising a larger payment when the target is reached. Conversely if aggregate default intensity is expected to deteriorate (λ_3 low relative to λ_2), the downsizing triggers a large drop in the bank's size-adjusted rent. There is no need to provide the bank with high-powered incentives and the size of probation is smaller.

3.4 Four loans: Backward expansion

The results for $i = 4$ are similar, except for the possible extension of the target beyond the end of the second interval $b_4 + b_3 + b_2$. With

$$w_4^2(\theta) = B_4^2 + A_4^2\theta + \phi_4^2(\theta) - C_4^2 \frac{\Lambda_4^p}{\lambda_4} \left(\frac{\Lambda_4^p + r\theta}{\Lambda_4^p} \right)^{\lambda_4/r},$$

the boundary $\overline{\theta}_4$ in that interval, if it is an interior solution, is again determined by the pasting conditions $\dot{w}_4^2 = -1$ and $\ddot{w}_4^2 = 0$ leading to

$$1 + \dot{w}_3^1(\theta) = \frac{r}{\lambda_4}. \tag{4}$$

The social value of performance is $1 + \dot{v}_3(u)$ one step ahead. The current relative cost of performance is r/λ_4. Whether or not expansion is warranted in state 4 depends on which interval one step ahead has a social value of performance equal to the current cost. If this happens when $v_3(u) = b_2 w_3^1(\theta)$, the first interval of state 3, condition (4) is met in the second interval of state 4. Investors want the best performers to be under threat of stochastic liquidation should a loss occur and there is no need to expand probation. If this happens when $v_3(u) = b_1 w_3^2(\theta)$, condition (4) cannot be met because the social value of performance is still high relative to its cost at $b_3 + b_2$. Investors want the best performers to fall in the probation interval of state 3 and this means expanding probation in state 4. In this still rather special case no backward *contraction* can take place because there is no interval to pick before stochastic liquidation in state 3. This need not be the case for higher order states with a large number of intervals. Thus looking forward, contraction is by one interval at most, but expansion can be sizable.

The intuition behind this result is the following. Suppose $i = 4$ is a state where "contagion" has spread with maximum individual risk ($\alpha_4 = \alpha_3$). Looking forward, the bank's size-adjusted reservation rent *grows* by 33% following default ($b_3/3 = 4/3\, b_4/4$). This creates risk shifting incentives when performance is poor. To mitigate those risks, investors design a long probation interval. As shown below, the size-adjusted target, and consequently the payment made, are higher. So the size-adjusted reservation rent is improved in the advent of default, but the punishment is also more severe if the bank shirks in the bliss state.

If on the contrary delinquencies are not likely when there are four loans ($\alpha_4 < \alpha_3$), the bank enjoys high rents in that state and is undermined by the downsizing. There is no need to shrink probation going forward. In the advent of default the bank has to wangle its way into *both* the stochastic liquidation and probation intervals of state 3. The punchline is that, looking forward, the cost of performance r/λ rises slowly during spells of constant individual risk and small cuts in the probation size hold the bank's rents in check. In contrast, the cost of performance dwindles during phases corresponding to a sharp worsening of credit risk and lump increases in the probation size help restore its incentives to monitor.

If $1 + \dot{v}_3(b_3 + b_2) > r/\lambda_4$, the smooth pasting condition for the boundary $\overline{\theta}_4$ in the *third* interval reads

$$1 + \dot{w}_3^2(\theta) = \frac{r}{\lambda_4}.$$

By construction $1+\dot{w}_3$ vanishes at $\theta = \overline{\theta}_3$ so $\overline{\theta}_4 < \overline{\theta}_3$ for positive r. In the advent of default in the bliss state, the bank finds itself *within* probation and does not get payments for some time. The targets consistent with the bank being paid in state $i = 4$ and $i = 3$ are

$$\begin{aligned}\gamma_4 &= b_4 + b_3 + b_2 + \overline{\theta}_4 b_1 \\ \gamma_3 &= b_3 + b_2 + \overline{\theta}_3 b_1,\end{aligned}$$

respectively. Since $b_4 \geq b_3 \geq b_2 \geq b_1$, the size-adjusted gap is minimized when all b_i are equal, $\overline{\theta}_4 = 0$ and $\overline{\theta}_3 = 1$, yielding

$$\frac{\gamma_4}{4} - \frac{\gamma_3}{3} \geq b_4 \left(\frac{1}{4} - \frac{1}{12} - \frac{1}{12} - \frac{1}{12} \right) = 0.$$

The bank loses when probation contracts, not only in absolute terms, but also in relative terms. This is not necessarily the case otherwise.

3.5 General case

We can now state the following.

Proposition 2 *Under Assumptions 1 to 3, the solution of the HJB system of equations*

$$\begin{aligned} & \max_{\delta_t(\cdot)\geq 0} && \left\{\dot{v}_i\left(ru+\lambda_i b_i-\delta_t\right)+i\mu-\delta_t-\lambda_i\left(v_i(u)-\theta v_{i-1}(u-b_i)\right)\right\}=0 \\ s.t. \quad & du+\delta_t\,dt &=& \; ru\,dt-b_i\wedge(u-b_{i-1})\left(\sum dN_t^i-\lambda_i\,dt\right)-b_{i-1}\left(dM_t-(1-\theta)\lambda_i\,dt\right) \\ & \theta &=& \; \frac{u-b_i}{b_{i-1}}\wedge 1 \end{aligned}$$

has maximal solutions $v_i(u)$ over $[b_i,\infty)$. The functions v_i are globally concave, continuously differentiable, with first positive slope and eventually slope -1 over $[\gamma_i,\infty)$, where

$$\gamma_i=\sum_{j=0}^{l(i)} b_{i-j}+\overline{\theta}_i b_{i-l(i)-1}, \qquad \overline{\theta}_i\in[0,1],$$

is the target rent in state i. On $[b_i, b_i+b_{i-1})$ there is stochastic liquidation given default with probability $1-\theta$. On $[b_i,\gamma_i)$ payment is differed. The cut-off rule $l(i)$ satisfies $l(i+1)\leq l(i)+1$, with $l(1)=l(2)=0$ and $l(3)=1$. The scale $\overline{\theta}_i$ is the probability of survival after $l(i)+1$ joint defaults in the bliss state, with $\overline{\theta}_{i+1}\leq\overline{\theta}_i$ if $l(i+1)=l(i)+1$ (strict inequality if $r>0$) and $\overline{\theta}_1=0$, $\overline{\theta}_2=1$. The cut-off rule and scale $\left(l(i),\overline{\theta}_i\right)$ are uniquely determined by the recursive conditions

$$1+\dot{w}_{i-1}^{l(i)}\left(\overline{\theta}_i\right)=\frac{r}{\lambda_i}.$$

In particular $l(i)=i-2$ and $\overline{\theta}_i=1$ if $r=0$.

The optimal risk prevention policy relies on two instruments: the prospect of future payments if there is no loss for some time (the carrot), and the risk of stochastic liquidation if there is a spell of poor performance (the stick). This history dependence is summarized by two variables: past downsizing, reflected in the number of loans outsanding $i = I - N$, and past performance, reflected in the bank's informational rent u. The minimum rent consistent with monitoring is b_i. Given track record $u \geq b_i$, it makes sense for investors to encourage the bank to improve its credentials before making payments. To keep the bank participating, they let its rent grow at a rate consistent with size and rate of impatience. Proposition 2 determines how far the target γ_i is away from b_i. Once the target is reached, the bank is paid.

Suppose there are i loans outstanding, ordered by the rank in which they default, i.e., number i is the first to default, $i-1$ the next and so on. (It does not matter which particular loans are chosen, since they are identical.) The bottom rent b_i is associated with a 100% probability of liquidation given default ($\theta = 0$). Suppose instead investors depart from the stochastic liquidation rule and commit not to liquidate the bank if loan i defaults. Incentive compatibility requires $u_i - b_i \geq u_{i-1}$ so $b_i + b_{i-1}$ is the minimum rent consistent with one exemption from liquidation ($l(i) = 1$). Likewise, if investors commit to exempt the bank from liquidation for up to $l(i)$ successive defaults, the bank's rent immediately jumps to $\sum_{j=0}^{l(i)} b_{i-j}$. Hence $l(i)$ can be interpreted as the *cut-off rule* associated with state i. It is the maximum number of joint defaults that the bank can withstand without fearing liquidation under the best track record.

Under the optimal plan the level of commitment is contingent on the bank's past performance. If $u \in [b_i, b_i + b_{i-1})$, commitment is granted with probability $\theta = (u - b_i)/b_{i-1}$. The utility range $[b_i, \gamma_i]$ can be broken into $l(i) + 1$ "buckets" of weakly decreasing size $b_{i-1}, \ldots, b_{i-l(i)}, b_{i-l(i)-1}$, the last being scaled down by $\overline{\theta}_i$. If u happens to be in the kth bucket, $k-1$ exemptions are granted and the kth is reneged with some probability. The exemption process is interrupted when u hits γ_i and the bank is paid. In the worst case scenario (which happens with probability zero), $l(i)$ defaults knock the bank in one stroke and its continuation utility collapses to $u = \gamma_i - \sum_{j=0}^{l(i)-1} b_{i-j} = b_{i-l(i)} + \overline{\theta}_i b_{i-l(i)-1}$. The scale factor $\overline{\theta}_i$ is simply the probability of survival faced by the bank following $l(i) + 1$ simultaneous defaults in the bliss state. An immediate default for an encore and the pool is disposed of, since $\theta = 0$ when $u = b_{i-l(i)-1}$.

The cut-off rule cannot be decremented by more than unit one looking forward. One special case[7] arises when $r = 0$ and $l(i) = i - 2$. Since it makes no sense to defer payments when the bank is infinitely patient, investors lose an instrument and are better off letting the bank cling to its target $\sum_{j \leq i} b_j$ anyway. There is no risk of private benefit diversion since it enjoys the highest possible rent. This may be very costly. Investors' value can actually be increased by assuming a *deterministic* cut-off rule. Such rule would trade off the disposal of valuable assets against the saving on monitoring costs ex ante. Introducing deterministic liquidation when $r \geq 0$ would not qualitatively change our results, as the recursive solution would simply start from a prespecified level.

An impatient bank is given a less ambitious target, but with time-varying size. If underlying individual risk is expected to remain constant for some time, aggregate risk is declining and the size of probation tapers off as losses unfold. This does not take the bank far away from target, since target and bank's utility are reduced jointly, but lowers payments in size-adjusted terms. If underlying risk is expected to deteriorate, probation can be reset to a higher size. This worsens the bank's position relative to target, but improves its prospects if lucky enough to earn its way out of trouble.

4 Implementation

The optimal contract can be implemented with securitization under realistic assumptions. We consider a true sale transaction, as we want the control rights to pass on to a third party, the issuing entity, which buys the pool of loans with the proceeds of the sale of asset-backed securities to outside investors.

Consider a bank originating a pool of I identical long term loans. The bank seeks to maximize its profits and, with given portfolio size, minimizes the amount of capital needed. Its program at time 0,

$$\begin{aligned} \max_{u \geq b_I} \quad & u - K \\ \text{s.t.} \quad K & \geq I - v_I(u), \end{aligned}$$

[7] The solution obtained by taking limits when $r \to 0$ is well-defined, with exponentials replacing power functions.

shows that, when the constraint binds, social surplus $S = u + v_I(u) - I$ is maximized, implying $u^* = \gamma_I$ and $v^* = v_I(\gamma_I)$. This of course assumes that the bank has enough funds to start with, namely $K = I - v^*$.

The bank initiates an ABS transaction by selling the portfolio to a bankruptcy-remote trust with gain on sale S over the principal balance I. The trust is willing to pay this premium because the anticipated payments from the arrangement below ensure that it breaks even. The sponsoring bank then hires a servicer to conduct due diligence on the borrowers. We consider only the relationship between the sponsoring bank/servicer on the one hand and the trust on the other hand,[8] leaving out further aspects concerning securitization, such as consulting with credit agencies or underwriting new securities to outside investors.

Individual default intensities are sometimes taken piecewise constant in practice. Consider a CDO whose attachment points track the changes in the actual distribution of individual risk under the risk-neutral probability. Alternatively, estimate individual risk in exogenously given tranches. As long as it is constant, the reservation utility $b_i = B/(\epsilon\alpha_i)$ is also fixed. We assume that the tranches cover the whole spectrum of losses. (More realistically a deterministic cut-off could be set at some lower end point; see above.) Under systemic risk, the more senior the tranche, the worse its default characteristics.

A tranche $[L, U]$ yields protection $[N - L]^+ - [N - U]^+$, where the difference between the attachment points, $U - L$, is the notional size of the tranche. It reimburses losses between L and U, if any. The protection embedded in a portfolio of tranches with "optimal" weights adjusted to the underlying reservation utilities,

$$P(N) = \sum_{[L,U]} b_{L,U} \left([N - L]^+ - [N - U]^+ \right),$$

is just the sum of weights b_j when j runs the gamut from $j = I$ to $j = I - N + 1$. From the previous section, b_i is the utility foregone by the bank after the nth-to-default loan ($n = I - i + 1$). The default-contingent exposure $P(N)$ rises from zero to $\sum_{j=1}^{I} b_j$ and its growth is driven by the bank's constant *risk shares*[9] b_i/μ in the tranches. Over the first-loss piece b_I/μ is close to one and the bank takes the brunt of the losses to

[8] In the model, the servicer is affiliated with the sponsor and the two are treated on a consolidated basis. The servicer is assured of the sponsor's profit and there is no agency problem between them.

[9] Because only monitored finance is viable, the gains from monitoring $\mu(\lambda_i(1+\epsilon) - \lambda_i)$ are always larger than the private benefits from shirking Bi. Thus $b_i = B/\epsilon\alpha_i < \mu$.

protect investors. Over the senior tranche, the bank is less exposed to default risk and a larger fraction of losses is passed through to investors. According to the optimal contract, the bank keeps sharing in the risk at a declining rate, until liquidation. This is in marked contrast with standard practice in structured finance, where the common retention mechanism is one for the first-loss piece and zero for all other tranches. One arrangement works as follows.

Proposition 3 *If individual risk is constant within tranches, the optimal risk prevention policy can be implemented with securitization:*

(i) Collateral $u^* = \gamma_I$ is withdrawn from the sale and posted in a reserve account managed by the trust;

(ii) The bank buys CDO tranches, weighted $b = B/(\epsilon\alpha)$, and waives its rights to the premium spreads;

(iii) The protection embedded in the tranches is assigned to the reserve account;

(iv) The servicing fee Bi/ϵ and accrued interests (rate r) are credited to the reserve account;

(v) The account balance is maintained between cap γ_i and floor b_i:

— Excess cash triggers payment to the bank;

— Overdrafts trigger stochastic liquidation: the trust makes up for the shortfall if the pool is rescued, seizes the account and settles outstanding CDOs if it is liquidated.

The trust incentivizes the bank by subordinating cashflows to its performance record. First, the sponsor guarantees the deal by pledging $u^* = K + S$ out of the proceeds of the sale and places the funds in a reserve account managed by the trust. Second, the bank writes protection by buying CDO tranches (CDS style[10]) to match the optimal declining risk shares, using the reserve account as cash collateral. It does not matter whether the buyers of protection are outside investors or the trust itself. What matters is that the premium flows generated by the credit enhancement do *not* accrue on the reserve account, lest the bank were considered as a simple arbitrageur operating in the credit derivative market. Third, the servicing and rent-preserving fees always remain at the top of the flow of funds, whether they are directly remitted to the bank or serve to replenish the reserve account. Earmarking a portion of the premium spreads allows

[10] A CDS style deal involves no payment at inception: the premiums flow in exchange for capital protection paid as and when credit events occur; cf. Chaplin (2005).

the trust to cater for the sponsor's impatience with the same instruments as those used for the synthetic compensation.

Finally, the trust monitors performance continuously by peeking at the bank's cash position within prescribed limits. Should the bank tread the stochastic liquidation interval, heightening solvency concerns, a "regulator" with full commitment is called for. Were then the balance to fall beyond floor, the regulator decides whether liquidation is warranted, perhaps on the basis of her superior information. The trust always stands in for the problem bank, either settling with a cash payment if the pool is kept afloat, or seizing the account and insulating the buyers of protection from counterparty risk if allowed to go under. Since the arrangement regulates liquidity as in the optimal incentive-compatible contract, the bank maximizes profits subject to its conducting due diligence on borrowers and the trust breaks even.

5 Policy implications

The cost of mortgage debt has increased dramatically in recent months. Ouside investors and overseas buyers have backed away following concerns about the US housing market and uncertainty about the involvement of the US government in the support of agency debt. The breakdown in the subprime mortgage market is due in some part to informational frictions between borrowers, lenders and other key players in the securitization process. While the paper doesn't deal directly with the current crisis — systemic risk is modelled at the individual bank level only and there is no interbank market or interdependencies between banks — it has noteworthy implications. The overall punchline is that what we see may be more a flaw of regulation than one of securitization.

One issue is whether the ability to securitize changes the risk profile of bank balance sheets in the first place. With on-balance sheet lending, banks are disciplined by a standard debt contract.[11] The optimality of a standard debt contract when effort is undertaken in the beginning follows from Innes (1990) and can be viewed as an application of the principle of the deductible which, as recalled by Franke and Krahnen (2008),

[11] One modern version of this view is that banks' incentives are reinforced by the illiquidity of loans and the fragility of demand deposits (Diamond and Rajan, 2003).

is the “magic” trick of incentive alignement familiar from insurance contracts. In this world, banks that originate bad loans bear the impact of losses up to a FLP and act as good delegated monitors. With the business model of securitization, however, informational frictions that arise from two-tier and even multi-tier agency relationships complicate the delegated monitoring problem. The risk of private benefit diversion from those committing their specific collection skills or administering the pool of assets becomes a real issue. One first implication of this paper is that when continuous monitoring is relevant the risk profile of bank balance sheets changes and incentive alignement can no longer be achieved by a standard debt contract. Ironically, complex structured instruments deemed to be at the “heart” of the credit market woes provide a good basis to pass risks on to third parties in good economic sense.

A second issue is whether securitization structures are suitably accounted for by Basel requirements. Acharya and Schnabl (2008) argue that sponsoring banks were able to call something as off-balance sheet, lower their capital charge, and thus operate at a higher leverage than regulators perceived. The prevailing view among analysts is indeed that excessive leverage built up by banks has lead them to lend “down the quality curve.”

One problem with the Securitization Framework concerns the treatment of second loss positions. Banks are able to include their exposures in a second loss position or better in the calculation of their risk weighted assets under relatively mild conditions.[12] The paper suggests in contrast that all securitization exposures provided by the sponsoring bank for credit enhancement should attract a deduction. The size of sponsoring banks' exposures to securitization tranches must decrease with their seniority, but theory gives no reason why the regulatory treatment of second loss positions should be discounted relative to that of the first-loss position. This is especially true for the most senior exposure, for which the Basel requirements above are waived altogether. According to the model, the most senior exposure is also risky because liquidation is possible before it starts suffering losses. Whether patient or not, a bank can also be subject to deterministic liquidation when less than a given fraction remains outstanding. Certainly no tranche can be securitized in

[12] Namely (i) the exposure is economically second loss position and the first loss position provides significant protection (ii) the credit risk is rated investment grade (iii) the credit risk is unrated and the bank does not retain or provide the first loss position.

that fraction. The fact that basis correlation can be found to be as high as one in the recent environment seems indicative of faulty system design.

Another problem is that banks are not constrained to retain any substantial part of the risk and maintain it over time. In a traditional securitization, a bank may exclude all assets from its risk-based capital calculations, provided it complies with operational requirements prescribing that the assets remain beyond its reach and that of its creditors. If the sponsoring bank does not retain any risk, the ownership is transferred and there is no capital charge. This is the worst of all worlds, since Basel II recognizes that the sponsoring bank may retain the "servicing rights to exposures" without it constituting "indirect control of the exposures" and so remain in the possession of hidden information concerning the pool of assets. One might argue that the price of a securitization transaction conveys information about the underlying quality of loans. But disclosure of the amount paid for the pool is not required for assets that are not securities, on the ground that such information is proprietary and in some instances not a meaningful concept; cf. Federal Register (2005, IIIB3c).

A third issue is whether prudential regulation plays its role in ensuring that banks engage in optimal CRT. Suppose that after funds have been raised from deposits and loans made, a bank engages in CRT without being committed to the optimal plan. It can hold fewer junior tranches and more senior tranches than necessary. In good states the bank receives high fees relative to the protection sold. It has a high utility and keeps monitoring. In bad states the fees may fall short of the protection sold. The promise-keeping constaint breaks down and the bank stops monitoring. The trust breaks even if this is factored in the pricing, but the bank increases its revenue by shifting losses to depositors. As pointed out by Chiesa (2008), prudential regulation may have a role in solving this commitment problem and restoring efficiency. Casual evidence cited in Franke and Krahnen (2008) shows that "the allocation of risks in securitization transactions is one of the well guarded secrecies of the industry"and that despite inconsistencies in empirical studies "the observed risk transfer is probably quite different from what theory predicts." The paper concurs with Franke and Krahnen (2007) that "the actual allocation of these tranches to investors in the economy is of particular relevance for bank supervisors."

A fourth issue is that many structures do not have mark to market prices, and banks essentially mark them to their advantage since they are compensated short-term with the very high coupon paid on the FLP[13] and take out the capital needed to bear the risk in the long term. This is a problem of incentives rather than of securitization per se. The bank is not entitled to receive the coupons generated by the protection it extends. More importantly, the results of the paper suggest that capital requirements alone cannot correct misaligned incentives, but that liquidity regulation may bring them back to the fold. A credit enhancement mechanism based on a proper allocation of CDOs subordinates the cash flows to overall performance, without prejudice of the servicing fees which remain at the top of the flow of funds. It is explicit, rather than based on back-up credit lines or other forms of implicit support which overwhelm bank liquidity in crisis times. It is prefunded with the proceeds of the sale, in the form of a reserve account managed by the trust, and thus resembles capital insurance in that protection is called for upon the occurrence of losses. It is subject to a regulatory charge, since the CDO premiums remain with the issuing trust, except for the fraction returned as rent-preserving fees to the sponsors.

It is often suggested that one of the main issues with regard to Basel II is its focus on individual banks. Given that banks will remain regulated at the individual level, regulators must include a measure of liquidity risk induced by correlation in individual risk measures. The charge for liquidity risk embedded in the optimal plan is based on loss intensities that can be calibrated from market imputs such as CDO tranche premiums. It can be seen as a tax prepaid by sponsors for the contingent support they receive as a result of their limited liability at the time of liquidation. When losses begin unfolding capital is automatically supplied by sponsoring banks and the tax is high. Only in case of liquidation capital is overwhelmingly supplied by the trust and the liquidity tax eventually eschewed.

[13] The cashflow "waterfall" implied by actual CDOs usually allocates loan income according to descending priority. Excess interest payments from the mortgage pool are paid to the equity tranche holder provided some conditions, such as the interest coverage or overcollateralization tests, are met. Such payments can arise in principle even when the equity tranche has been used up; cf. Chaplin (2005).

6 Conclusion

While the literature generally considers endogenous liquidation values with exogenously given contracts (Schleifer and Vishny, 1992), here we endogenize contracts with exogenously given liquidation values. Our starting point is that among the various sources of informational frictions moral hazard may be as important as adverse selection. Monitoring may consist either of screening borrowers to reduce the proportion of less creditworthy types ex ante, or of services tailored to the borrowers to minimize the probability of losses down the road. In the paper we deliberately play down the first aspect and emphasize the second. Continuous monitoring reduces defaults on bank loans just as continuous testing of students reduces the probability of failure.[14] Placed in the context of securitization, this means that one of the key frictions that may have caused the subprime crisis is moral hazard between sponsors and servicers on the one hand and investors and their trustees on the other. As should be clear from several references in the literature, the definition of "servicer" does not only encompass the collection of the pool assets but also the maintenance and allocation of the pool itself, functions that are often referred to as "administration."

The model finds a role for supervision to the extent that losses are not permitted to exceed a prespecified cut-off rule. Servicers would prefer to keep the loans on their books for as long as possible, as this would increase the income they receive from the portfolio, and should be constrained in the amount of time they are allowed to operate. Likewise, sponsors have an incentive to tilt their risk sharing towards retaining too much senior risk and too little junior risk, and due diligence conducted by supervisors may help prevent that. The model fits in relatively well with current recommendations that the bankrupty code should be amended to allow for regulatory intervention ahead of bank insolvency. But it strongly suggests that market discipline might be better imposed by a well-designed credit enhancement scheme based on tranching than simply outsourced to regulatory supervision.

[14] I am indebted to Robert Krainer (U. of Wisconsin) for the analogy.

7 References

Acharya, Viral and Phillip Schnabl, 2008, "How banks played the leverage game," White Paper in *Repairing the US Financial Architecture*, NYU Stern School of Business, edited by Viral Acharya and Matthew Richardson.

Arping, Stefan, 2004, "Credit protection and lending relationships," Working Paper. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=318181.

Ashcraft, Adam and Til Schuermann, 2008, "Understanding the securitization of subprime mortgage credit," Federal Reserve Bank of New York Staff Reports, no. 318.

Berndt, Antje and Anurag Gupta, 2008, "Moral hazard and adverse selection in the originate-to-distribute model of bank credit." Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1290312.

Biais, Bruno, Thomas Mariotti, Guillaume Plantin and Jean-Charles Rochet, 2007, "Dynamic security design: Convergence to continuous time and asset pricing implications," *Review of Economic Studies*, 74, 345–390.

Biais, Bruno, Thomas Mariotti, Jean-Charles Rochet and Stéphane Villeneuve, 2009, "Large risks, limited liability and dynamic moral hazard," Mimeo, Toulouse School of Economics.

Cantor, Richard and Jian Hu, 2007, "Deal sponsor and credit risk of asset-backed and mortgage-backed securities," Moody's Investors Service, Working Paper.

Chiesa Gabriella, 2008, "Optimal credit risk transfer, monitored finance, and banks," *Journal of Financial Intermediation*, 17, 464–477.

Chaplin, Geoff, 2005, "*Credit Derivatives*," Wiley Finance.

Coval, Joshua, Jakub Jurek and Erik Stafford, "Economic catastrophe bonds," *American Economic Review*, forthcoming.

DeMarzo, Peter, 2005, "The pooling and tranching of securities: A model of informed intermediation," *Review of Financial Studies*, 18, 1–35.

DeMarzo, Peter and Darrell Duffie, 1999, "A liquidity-based model of security design," *Econometrica*, 67(1), 65–99.

DeMarzo, Peter and Michael Fishman, 2007a, " Agency and optimal investment dynamics," *The Review of Financial Studies*, 20, 151–189.

DeMarzo, Peter and Michael Fishman, 2007b, " Optimal long-term financial contracting," *The Review of Financial Studies*, 20, 2079–2128.

DeMarzo, Peter and Yuliy Sannikov, 2006, "Optimal security design and dynamic capital structure in a continuous-time agency model," *Journal of Finance*, 61, 2681–2724.

Diamond, Douglas and Raghuram Rajan, 2001, "Liquidity risk, liquidity creation and financial fragility: A theory of banking," *Journal of Political Economy*, 109, 287–327.

Dichev, Ilia and Douglas Skinner, 2001, "Large-sample evidence on the debt covenant hypothesis." Available at SSRN: http://ssrn.com/abstract=275174.

Drucker, Steven and Manju Puri, 2007, "On loan sales, loan contracting and lending relationships," BIS Working Papers, No 255.

Federal Register, 2005, 17 CFR Parts 210, 228, "Asset-Backed Securities; Final Rule, " Securities aand Exchange Commission.

Duffee, Gregory and Chunsheng Zhou, 2001, "Credit derivatives in banking: Useful tools for managing risks?" *Journal of Monetary Economics*, 48(1), 25–54.

Duffie, Darrell and Nicolae Gârleanu, 2001, "Risk and valuation of collateralized debt obligations," *Financial Analysts Journal*, 57(1), 41–62.

Duffie, Darrell, 2008, "Innovations in credit risk transfer: Implications for financial stability," BIS Working Papers, No 255.

Fender, Ingo and Janet Mitchell, 2008, "Incentives and tranche retention in securitisation: A screening model," Working Paper.

Franke, Günter and Jan Krahnen, 2007, "Default risk sharing between banks and markets: the contribution of collateralized debt obligations," in *The Risks of Financial Institutions*, ed. by M. Carey and R. Stulz (NBER book), 603–634.

Franke, Günter and Jan Krahnen, 2008, "The future of securitization," Working Paper. Available at

SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1284989

Gan, Yingjin and Christopher Mayer, 2006, 'Agency conflicts, asset substitution, and securitization," NBER Working Paper No 12359.

Holmström, Bengt and Jean Tirole, 1998, "Private and public supply of liquidity," *Journal of Political Economy*, 106(1), 1–40.

Innes, Robert, 1990, "Limited liability and incentive contracting with ex-ante action choices, *Journal of Economic Theory*, 52(1) 45–67.

Karlin, Samuel and Howard Taylor, 1975, "*A First Course in Stochastic Processes*," Second Edition, Academic Press.

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru and Vikrant Vig, 2008, "Did securitization lead to lax screening? Evidence from subprime loans." Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1093137.

Laurent, Jean-Paul, Areski Cousin and Jean-David Fermanian, 2007, "Hedging default risks of CDOs in Markovian contagion models." Available at DefaultRisk: http://www.defaultrisk.com/pp_cdo_26.htm.

Norden, Lars and Martin Weber, 2008, "Checking account information and credit risk of bank borrowers." Available at SSRN: http://ssrn.com/abstract=1101548.

Parlour, Christine and Andrew Winton, 2008, "Laying off credit risk: Loan sales versus credit default swaps," Working Paper.

Plantin, Guillaume, 2004, "Tranching," Working Paper, London School of Economics.

Sannilkov, Yuliy, 2005, "Agency problems, screening and incresing credit lines," Mimeo, University of California, Berkeley.

Sannikov, Yuliy, 2008, "A contnuous-time version of the principal-agent problem," *Review of Economic Studies*, 75, 957–984.

Schleifer, Andrei, and Robert Vishny, 1992, "Liquidation values and debt capacities: A market equilibrium approach," *The Journal of Finance*, 47(4), 1343–1366.

8 Proof of Proposition 3

From the bank's integrated promise-keeping constraint (2) along the optimal path, we know that for all $t \leq \tau$

$$u_t = u^* + \int_0^t \left(ru_s + \frac{Bi}{\epsilon} \right) 1_{\{u_s < \gamma_i\}}\, ds - \int_0^t b_i \wedge (u - b_{i-1}) \sum_j dN_s^j - \int_0^t b_{i-1}\, dM_s,$$

where $i = I - \sum_j N_t^j$. By construction, the protection sold by the bank is

$$P(N_t) = \int_0^t b_i \sum_j dN_s^j + \left(\sum_{j<i} b_j \right) M_{t\wedge\tau}$$

since at $t = \tau$ the default count jumps from $\sum_j N_\tau^j$ to I. Thus

$$u_t + P(N_t) = u^* + \int_0^t \left(ru_s + \frac{Bi}{\epsilon} \right) 1_{\{u_s < \gamma_i\}}\, ds + \xi_t + \left(\sum_{j<i} b_j \right) M_{t\wedge\tau} \tag{5}$$

where the martingale

$$\xi_t = \int_0^t [b_i + b_{i-1} - u]^+ \sum_j dN_s^j - \int_0^t b_{i-1}\, dM_s$$

is the trust's cumulated cost resulting from intervention after stochastic liquidation. With probability θ, the bank is rescued and the trust pays $\Delta\xi = b_i + b_{i-1} - u$. With probability $1-\theta$ the bank is closed and the trust wins $-\Delta\xi = u - b_i$. Evaluating (5) at $t = \tau$ with $u_\tau = 0$ and $N_\tau = I$, we get

$$\sum_{j\leq I} b_j = u^* + \int_0^\tau \left(ru_s + \frac{Bi}{\epsilon} \right) 1_{\{u_s < \gamma_i\}}\, ds + \xi_\tau + \sum_{j<i^*} b_j \tag{6}$$

where $i^* = I - \sum_j N_\tau^j$ is the portfolio size at liquidation.

The bank maximizes its profit since by construction

$$\begin{aligned} u^* &= E \int_0^\tau e^{-rt} \left(r\gamma_i + \frac{Bi}{\epsilon} \right) 1_{\{u_t = \gamma_i\}}\, dt \\ &= E \int_0^\tau e^{-rt} \delta_t\, dt. \end{aligned}$$

The trust's costs and benefits in the course of the relationship are as follows

	Cost	Benefit
$t = 0$	$u^* + v^*$	—
$(t, t + dt)$	$\left(ru_t + \frac{Bi}{\epsilon}\right) 1_{\{u_t < \gamma_i\}}$	$i\mu - \delta_t + \Sigma_t$
Liquidation	$\xi_\tau + \sum_{j<i^*} b_j$	—

where Σ_t is the premium flow from the CDO tranches. From (6), the overall cost is deterministic and equal to $v^* + \sum_{j \leq I} b_j$. But

$$\begin{aligned} v^* &= E \int_0^\tau (i\mu - \delta_t)\, dt \\ \sum_{j \leq I} b_j &= E \int_0^\tau \Sigma_t\, dt, \end{aligned}$$

the first equality by design, the second by arbitrage. The trust breaks even.