Bank monitoring incentives and optimal CDOs

Henri Pagès*
Banque de France, DGEI

March 27, 2009

Abstract

The paper examines a delegated monitoring problem between investors and a bank holding a portfolio of correlated balloon loans displaying "contagion." Moral hazard prevents the bank from monitoring continuously unless it is compensated with the right incentive-compatible contract. Under the Monotone-Likelihood-Ratio Property (MLRP), the asset pool is liquidated when losses exceed a 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. Securitization can approximate the optimal contract. The sponsor provides the trust with credit enhancement in the form of a weighted portfolio of collateralized debt obligations which is funded by the proceeds of the sale and extends into the cut-off rule. In compensation the trust pays servicing fees as well as rent-preserving fees if the sponsor has a high discount rate. Rather than being detrimental, well-designed securitization seems an effective and convenient means of implementing the second best.

Keywords: Credit risk transfer, Default Risk, Contagion

JEL classification: G21, G28, G32

*Banque de France, DGEI, 41-1431, *E-mail address:* henri.pages@banque-france.fr. I thank Viral Acharya, Gilles Chemla, Monique Jeanblanc, Jean-Charles Rochet as well as participants in an internal seminar for helpful and encouraging comments. All errors are mine.

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 — "forward 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.

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 is improved 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 forward 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 focuses on an individual bank, the systemic risk contribution of the portfolio is handled with a model of "contagion." The model is Markovian, with individual default intensities depending on the number of non-performing loans already observed. Consistently with the empirical evidence documented by Laurent at al. (2007), contagious dependence between defaults is introduced by assuming that the smaller the principal balance of the portfolio (the larger the number of non-performing loans), the higher the aggregate default intensity. This imperfect correlation between default times undermines the bank's ability to diversify its credit risk and makes senior tranches comparable to "economic catastrophe bonds,"[1] i.e., low in risk overall, but likely to be wiped out if the risk materializes.

To simplify the exposition we consider a portfolio of identical "balloon" mortgage loans yielding a return only at the final date. We show that the optimal compensation entails both continuous fees and a default-dependent payoff received when mortgages convert into cash. Under a maximum likelihood ratio property (MLRP), the bank absorbs losses at a declining rate, up to a threshold defining the highest permissible level for losses. Otherwise, the portfolio is liquidated and neither the bank nor investors receive anything. This is achieved through a cut-off rule and a convex payoff schedule specifying how much of the portfolio return accrues to the bank when the critical level is not reached. The intuition is the following. In the beginning the aggregate default intensity is low and it is difficult to disentangle a bank which monitors from one which does not. The extra rent for monitoring one more loan must be high. So the bank has initially a high rent and bears the brunt of losses to have high-powered incentives to monitor. In contrast, when a substantial

[1] Coval et al., AER, forthcoming.

fraction of the portfolio is exhausted after bouts of contagious defaults, the higher aggregate default intensity and the threat of termination imply that the bank is eager to monitor. In bad states the monitoring bank has lost much of its initial rent and must be protected against losses associated with financial distress.

Although subordinating cash flows to the performance of the portfolio helps align the interests of the bank and investors, the latter must advance fees to the former during the transaction. The reason is twofold. The bank must be shielded against the incidence of losses over which it does not have control, as this would erode its payoff even though monitoring actually takes place. This cost-of-servicing fee is a flat percentage of the outstanding portolio. The bank may also be more impatient than investors, implying that more fees must be paid out to keep its current informational rent growing. This rent-preserving fee contributes towards strengthening the cut-off rule and the capital requirement to restore investors' break-even condition.

These findings have two implications. First, there is a limit to the extent of default risk that can be transferred to outside investors. Liquidation can be useful if high cumulated losses signal that the bank may not have behaved properly. It is inefficient ex post since it involves the disposal of valuable assets, but allows saving on monitoring costs ex ante. Writing off the portfolio when less than a given fraction remains outstanding is a sensible option when losses become contagious. Second, the contractual mechanism to influence the bank's choice of costly effort is not retention of a first-loss piece (FLP). The size of expected losses can be significant in comparison with the FLP (6% for subprime in late 2006) and if bad performance exhausts the value of the equity piece, the bank has an incentive to save on servicing expenses or indulge in asset substitution. The model actually provides a clean answer in the light of the amendment by the European Commission (2008) to the European Capital requirement directive proposing that banks have more "skin in the game." Optimal credit risk transfer is supported by a structure of partial credit enhancements that taper off along the permissible spectrum of losses and are backed by the bank's claim on future returns.

The theoretical model specifies that the bank has sole rights over the cash flows of its portfolio. In the implementation section we show how the optimal contract can be approximated through a true sale transaction. The idea is that since securitization is based on a portfolio rather than individual loans, banks can avail themselves of multiple classes of asset-backed securities to synthetize the desired protection.

One arrangement works as follows. The sponsoring bank sells its portfolio to a trust and allocates the administering function to an affiliated servicer. (In our model, the servicer must be affiliated with the sponsor because the two are treated on a consolidated basis. The sponsor is assured of the servicer's profit, and there is no agency problem between them.) The sponsor pledges protection to the trust by buying a portfolio of collateralized debt obligations, in credit-linked note (CLN) form, out of the proceeds of the sale. The various tranches cover loan losses up to the cut-off rule, with their respective nominal adjusted to fit the optimal structure of credit enhancements. Both servicing fees paid by the trust are at the top of the flow of funds. Part of the CLN premiums actually hedge the rent-preserving fee if the sponsor has a higher discount rate than investors. In any case the remainder stays with the trust as a liquidity charge the sponsor has to pay for the systemic risk borne by the trust. The residual value of the CLNs is the default-dependent compensation for the sponsor, and the trust breaks even. The result shows that, with forward 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 is related to several 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 it is 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. There is a distinct literature on the role of implicit recourse in securitization, which is related to the extent that the protection extended by originators should bolster their incentives to invest in quality servicing. In Gorton and Souleles (2005), the yields for credit card ABS reflect market beliefs about sponsors' ability to meet their commitments. Their work echoes Higgins and Mason (2004), who find that credit card originators appear to provide recourse to assuage concerns about the quality of their 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 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) extends 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 forward monitoring.

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

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 (2007) 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 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.

2 The model

Consider an economy populated by a bank and investors. The former has sole rights to the returns of specific investment opportunities at date 0, and the latter supply liquidity competitively. There is universal risk neutrality. The bank's opportunity set consists of I unit loans that are identically distributed, imperfectly correlated, and yield a risky return $R > 1$ at date 1. One can view them as "balloon" subprime mortgages[3]

[3] A balloon mortgage does not fully amortize over the term of the loan and requires a large final (balloon) payment. The proportion of balloon mortage contracts jumped substantially in 2006, accounting for 25 percent of all originations in the US securitized subprime market (Demyanyk and Van Hemert, 2008).

to firms or individuals, the default risk of which has some systematic component.[4] 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 will instead emphasize 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 *forward 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. We focus on contracts that implement monitoring as long as the bank is a going concern. 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 outstanding portfolio. We show that, if investors can commit to liquidate loans before maturity, they will ensure that the bank is diligent by winding it down when losses exceed a given threshold.

[4] The model has some similar features to Holmström and Tirole (1998), with correlated defaults among loans assuming the role of the interim liquidity shock.

2.1 Forward monitoring

Let N_t, with $0 \leq N_t \leq I$ and $t \in [0, 1]$, be the number of loan defaults in the portfolio at time t. All loans are a priori identical and it does not matter which particular loans default. Initially $N_0 = 0$ and if $N_t = I$ the bank has reached the final (absorbing) state where all is lost. The information is that generated by default times, with bank and investors only observing how many loans have failed up to time t. The process N_t is taken to be time-homogeneous and Markovian (Karlin and Taylor, 1975). The Markovian credit dynamics is flexible enough to allow calibration on any loss distributions at a given horizon,[5] as for example that specified in Li's benchmark model of correlated defaults (Li, 2000).

More specifically, we consider a model in which current monitoring effort only affects the instantaneous probability of loss. The bank can either "behave" or "shirk." With forward monitoring the risk-neutral aggregate loss intensity is $\lambda(N_t)$ in state N_t. It is related to individual risk-intensities by

$$\lambda(N_t) = (I - N_t)\ \gamma(N_t),$$

where $I - N_t$ is the number of loans outstanding at time t and the individual risk-intensities $\gamma_i(\cdot)$, $1 \leq i \leq I$, are equal across loans. (Independence between defaults corresponds to the case where the γ's are also fixed.) If the bank chooses not to monitor, it enjoys a private benefit $B(N_t)\,dt = B\,(I - N_t)\ dt$ between t and $t + dt$, in which case the aggregate loss intensity, $\lambda'(N_t)$, is higher than what it would be under monitoring. We assume that the effectiveness of monitoring is uniform in N, so that $\lambda'(N) = (1+\epsilon)\lambda(N)$. Both distribution of losses with or without monitoring can be recovered from the aggregate intensities λ_0, λ_1, $\ldots$, λ_{I-1} and ϵ.

Let $u(t, N)$ be the contractual time t continuation utility for the bank if there are N defaults in the portfolio at that time. Up to second-order terms[6] the bank's expected continuation utility at time $t + dt$ is

$$u(t, N_t)(1 - \lambda(N_t)\,dt + u(t, N_t + 1)\lambda(N_t)\,dt$$

[5] Calibrating the model at different points in time would require dropping the time-homogeneous assumption.

[6] In this homogeneous Markovian contagion framework, an important assumption is that only single defaults occur, i.e., N_{t+dt} is either N_t with probability $1 - \lambda\,dt$ or $N_t + 1$ with probability $\lambda\,dt$.

if it monitors and

$$u(t, N_t)(1 - \lambda'(N_t)\, dt) + u(t, N_t + 1)\lambda'(N_t)\, dt + B(I - N_t)\, dt$$

if it does not. We assume that conditional upon the ith loan defaulting the next state is $N_t = i + 1$, i.e., no partial liquidation is possible. Consequently, the bank will choose to monitor at time t if and only if

$$u(t, N) - u(t, N + 1) \geq \frac{B(I - N)}{\epsilon\lambda(N)}. \tag{1}$$

This is the usual incentive compatibility constraint. The quantity on the right-hand side can be interpreted as the non-pledgeabe income necessary to keep the bank diligent in state N given its pay-off in state $N + 1$. Note that it is not the *full* non-pledgeable income in state N, but the *extra* informational rent associated with monitoring one more unit in the portfolio.

2.2 Optimal policy

Consider the problem faced by risk-neutral investors entrusting their funds to the bank at time 0. The proceeds from investment are recouped only at the final date (balloon payments). The optimal contract may contain a termination call when N_t exceeds some prespecified level $i^*(t) < I$, in which case the final payoff from the portfolio is forgone. In addition to the bank's continuation utility $u(t, N) = u_N(t)$, the contract specifies an instantaneous fee $\delta(t, N) = \delta_N(t) \geq 0$ per unit time.

Investors are subject to a promise-keeping constraint ensuring that the bank is willing to participate at all times:

$$\dot{u}_i(t) + \delta_i(t) - \lambda_i\left(u_i(t) - u_{i+1}(t)\right) = r u_i(t), \tag{2}$$

where $r \geq 0$ is the bank's discount rate. The interest rate is assumed constant and normalized to zero. Equation (2) states that the bank's expected capital gains and dividends must be equal to the risk-adjusted discount rate. Dividends include the non-negative fee $\delta_i(t)$ and a negative dividend $u_{i+1}(t) - u_i(t)$ equal to the expected loss in continuation utility resulting from default risk at that time (with intensity λ_i).

Following DeMarzo and Fishman (2006), let $v_i(t, u)$ be investors' continuation utility in state $N_t = i$

when the bank's continuation utility is u. The optimal control problem leads to the following recursive system of dynamic programming equations indexed by $i = 0, 1, \ldots, I$:

$$\begin{aligned}
&\max_{\delta_i(\cdot)\geq 0} && \left\{ \frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial u}\dot{u}_i + \lambda_i \left(v_{i+1}(t, u_{i+1}(t)) - v_i(t, u_i(t))\right) - \delta_i \right\} = 0 \\
&\text{s.t.} && \dot{u}_i + \delta_i(t) - \lambda_i \left(u_i(t) - u_{i+1}(t)\right) = ru_i(t) \\
& && u_i(t) - u_{i+1}(t) \geq B(I-i)/\epsilon\lambda_i
\end{aligned}$$

with limit conditions

$$\begin{aligned}
u_i(t) &= v_i(t,u) = 0 \qquad \text{for all } (t,i,u) \text{ such that } i > i^*(t) \\
v_i(1,u) &= R(I-i) - u \qquad \text{for all } u \text{ and } i \leq i^*(1)
\end{aligned}$$

Proposition 1 *Suppose the bank operates in state $N_t = i$. Then its continuation utility and fee are defined recursively by*

$$\begin{aligned}
u_i(t) &= u_i = u_{i+1} + \frac{B\,(I-i)}{\epsilon\lambda_i} \\
\delta_i(t) &= \delta_i = B(I-i)/\epsilon + ru_i.
\end{aligned}$$

Investors' continuation utility is given by

$$v_i(t, u_i) = w_i(t) - u_i$$

where

$$\dot{w}_i(t) = \lambda_i \left(w_i(t) - w_{i+1}(t)\right) + ru_i$$

with terminal condition $w_i(1) = R(I-i)$.

The bank's continuation utility is set at the minimum level consistent with its incentives to monitor.

Monitoring is costly in state i but allows the bank to enjoy u_i rather than u_{i+1} with higher probability. The private opportunity cost $B(I-i)$ must equal the private benefit $\epsilon\lambda_i(u_i - u_{i+1})$. Maintaining u_i constant, however, requires paying a fee both because the bank must be compensated for the risk of loss and its discount rate r is above the riskless rate. The level set for δ_i ensures that $\dot{u}_i = 0$ in the promise-keeping equation (2). An alternative would be to pay less fees while letting the bank's utility grow above the minimum level u_i. This option however would entail a higher payoff at $t = 1$ when $r > 0$. Postponing payments only makes investors worse off as the bank is more impatient. The bank receives u_i as final payoff if it operates in state i. Investors are left with the net return from the portfolio $v_i(1, u_i) = R(I-i) - u_i$.

In state i the size of the bank remains constant. With constant returns to scale, investors' continuation utility $v_i(t, u) = w_i(t) - u$ depends linearly on the bank's continuation utility. Investors thus maximize their utility by setting u as low as possible and get $v_i(t, u_i) = w_i(t) - u_i$. The joint continuous utility, $w_i(t)$, yields a capital gain both because the current state may terminate and because the bank is impatient. This, with the terminal condition, completely determines continuation utilities in operating state i at all times.

2.3 Cut-off rule

The bank can be liquidated before the portfolio is depleted. The policy prescription is that the bank should be closed when N_t exceeds some threshold $i^*(t) < I$. The proposition shows that this critical level, if there is one, does not depend on time. For suppose that the bank is liquidated when the default count N exits the threshold i^* at time t, but not after $s > t$. Since the bank's continuation utility should remain constant between t and s, $u_{i^*}(s) = B(I - i^*)/\epsilon$ is the incentive-compatible level consistent with i^* being the last state. This is inconsistent with the cut-off rule being set at a higher level than i^*.

The contract must determine the optimal theshold i^*. Liquidating the portfolio precludes the option of collecting loans but saves on monitoring costs. The outside option of a lender operating at the threshold of liquidation is low, because everything is lost when a single loan fails. Looking backwards, this implies that the lender's continuation utility is lower in all states preceding liquidation. It might not make sense to write off loans to save on monitoring costs if the probability of collecting them were high. But with contagion, the aggregate intensity of default increases, while the unit opportunity cost of monitoring remains constant.

We proceed under the assumption that the Markovian contagion model with fixed parameters satisfies the monotone likelihood ratio property, and show that liquidation can improve the bank's incentives to monitor. That MLRP should be satisfied in the general case is by no means a forgone conclusion. Suppose for example that there are four loans in the portfolio, with $\lambda_0 > \lambda_3$. Such is the case under independence, since the intensity of default is then proportional to the number of assets outstanding and $\lambda_0 = 4\lambda_3$. Under uniform effectiveness of monitoring, states with high likelihood ratios are associated with high rates of change in their probability of occurring; cf. proof of Proposition 2. Will $q_3 = P(N_t = 3)$ grow at faster rate than $q_2 = P(N_t = 2)$? The answer is no. Although q_3 must initially grow as more delinquencies eventually lead to a single loan, the last state is resilient to the extent that λ_3 is small. In contrast, defaults are much more likely when there are three or four loans in the portfolio, and this causes q_2 to increase faster.

In a recent paper, Laurent, Cousin and Fermanian (2007) show that, when the time-homogeneous Markov model above is calibrated on the loss distribution generated either by a Gaussian copula (Li 's model of correlated default) or by market inputs such as quoted CDO tranche premiums, the fitted loss intensities increase with the number of defaults. They change almost linearly under the Gaussian copula assumption, and even more quickly using a base correlation curve for the iTraxx. This gives credence to the contagion hypothesis, according to which the default intensity of the Nth-to-default loan increases with N. We conjecture that under the contagion hypothesis, the time-homogeneous Markovian model satisfies MLRP.[7]

Proposition 2 *Let q_i be the probability that $N_1 = i$ for $0 \leq i < I$ and τ_i the exit time from $N_t = i$. Under MLRP, there exists $i^* < I$ such that the bank is closed when losses exceed i^* if $B/\epsilon\lambda_0 \left(1 + rE\tau_0 \wedge 1\right) \leq Rq_0$. The cut-off rule i^* is defined as the highest i such that*

$$\frac{B}{\epsilon\lambda_i}\left(1 + rE\tau_i \wedge 1\right) \leq Rq_i. \tag{3}$$

[7] We can show that MLRP is satisfied by the time-homogeneous Markovian contagion model up to rank $I - 1$ when there are no more than four loans when $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{I-1}$. We conjecture that the property remains valid irrespective of the number of names in the portfolio and up to the last rank.

The bank's and investors' continuation utilities at time 0 are respectively

$$\begin{aligned} u_0 &= \sum_{i=0}^{i^*} \frac{B(I-i)}{\epsilon\lambda_i} \\ v_0 &= \sum_{i=0}^{i^*} (I-i)\left(Rq_i - \frac{rB}{\epsilon\lambda_i} E\tau_i \wedge 1)\right) - u_0. \end{aligned}$$

Suppose first $r = 0$. The cut-off rule i^* guarantees that the bank operates as long as $B/\epsilon\lambda_i \leq Rq_i$. In each operating state, the extra rent consistent with diligent monitoring is less than the revenue generated. It turns out that, under MLRP, the probability that the ith loan defauts at date 1, $\lambda_i q_i$, is first increasing with i and eventually vanishing. (In any event $\lambda_I q_I = 0$ since I is absorbing.) Thus the set of ranks satisfying $B/R\epsilon \leq \lambda_i q_i$ is bounded above by some i^*.

When the discount factor is positive, dividends must be bulked out to compensate the bank for higher impatience. The rent associated with state i is $u_i - u_{i+1}$. The rent-preserving fee, $r(u_i - u_{i+1}) = rB(I-i)/\epsilon\lambda_i$, is paid until the bank exits the given state, at time τ_i, or until maturity, whichever is smaller. The effective cost incurred per unit loan is that on the left-hand-side of (3). It now increases with the number of defaults rather than being constant since the higher i, the longer the exit time from state i. The cut-off rule i^* is shortened when the discount rate is positive. The assumption stated at the beginning of the proposition ensures that the expected cost does not exceed expected revenue at inception.

The choice of i^* can be interpreted as the outcome of second-best risk sharing. At date 0 the bank can borrow up to $v_0 = w_0 - u_0$ from investors, the social continuation value of the portfolio minus the non-pledgeable income for monitoring, and is allowed to operate as long as there are at least $I - i^*$ loans in the portfolio. If there are few defaults till maturity, $v(t)$ remains positive and investors win. If contagion depletes the portfolio before maturity, $v(t)$ becomes negative and investors loose. Take for example $I = 2$ with $\lambda_1 e^{-\lambda_1} < B/R\epsilon < \lambda_0 e^{-\lambda_0}$ and $r = 0$. In that case $\lambda_0 q_0 < \lambda_1 q_1$ and Proposition 2 implies that $i^* = 1$ (the bank operates until both loans have defaulted). But $v_1(0) = Re^{-\lambda_1} - B/\epsilon\lambda_1 < 0$ so if the first default occurs shortly after date 0 investors' continuation utility is negative. (The expected net payoff at maturity $e^{-\lambda_1}(R - u_1)$ falls short of the present value of fees $\left(1 - e^{-\lambda_1}\right) u_1$.)

If $N = i \leq i^*$ at some point in time, the bank enjoys u_i, the sum of rents $\beta_j = B(I-i)/\epsilon\lambda_j$ associated with all non-defaulted loans $i \leq j \leq i^*$. The default-sensitive payoff $u(N) = \sum_{i=0}^{i^*} u_i 1_{\{N=i\}}$ has slope $-\beta_i$ at $N = i$. With a granular portfolio,[8] $R > \beta_0$. Moreover $\beta_0 > \cdots > \beta_{i^*}$ under the contagion hypothesis, because there are less loans to monitor and a higher aggregate default intensity as the number of defaults edges up. The share of rents in the return of the ith-to-default loan, β_i/R, can be interpreted as the bank's *long term risk share* in state i. If it is close to one, the bank takes the brunt of the loss to protect investors. If it is close to zero, the bank is weakly exposed to default risk and a larger fraction of the loss is passed through to investors. According to the optimal contract, the bank keeps sharing in the risk at a declining rate, until the portfolio is liquidated. It has high-powered incentives in the beginning but is more protected in later years if N becomes large. The decline is not directly motivated by agency problems with other parties, but by considerations about credit risk. This is in marked contrast with standard practice in structured finance, where the long term risk share implied by the common retention mechanism is one over R for the first-loss piece and zero for all other tranches.

3 Implementation

In the former section the arrangement that the bank and investors come to was left undetermined. We now show how the optimal contract can be approximated with securitization, assuming piecewise constant aggregate default intensities. 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 balloon loans. The bank initiates an asset-backed securities 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 sponsor then hires a servicer to conduct due diligence

[8] For large I, $\beta_0 < R$ under the assumption of Proprosition 2 since $\beta_0 = BI/\epsilon\lambda_0 \leq Rq_0 I$ and $q_0 I \to 0$ when $I \to \infty$ in any reasonable model of credit risk (such as Li's model).

on the borrowers. We consider only the relationship between the sponsor and the servicer on the one hand and the trust on the other hand, leaving out further aspects concerning securitization, such as consulting with credit agencies or underwriting new securities to outside investors.

Replication of the final pay-off $u(N)$ could be done exactly with Nth-to-default baskets. Indeed, N2D swaps would allow the sponsor to credit-enhance the deal by giving the trust the option to sell back the ith-to-default loan for price β_i. However, such instruments are small portfolio products and trading beyond the first-to-default swaps is rarely seen in practice. Instead, we rely on a super-replication strategy and implement another, more generous, incentive-compatible payoff $\widetilde{u}(N)$ under the requirement that the sponsor's long term risk share be constant over exogenously given tranches. A natural tranching arises if aggregate default intensities cannot be precisely estimated. We assume that there exists a partition L of $[0, i^* + 1]$ with attachment points $l_0 = 0 < l_1 < \cdots < l_n = i^* + 1$ such that the aggregate default intensity λ_{l_k} prevails in the tranche $[l_k, l_{k+1})$. Under systemic risk, the more senior the tranche, the worse its default characteristics.

With constant risk shares in each tranche the incentive compatibility constraints

$$\widetilde{u}_i - \widetilde{u}_{i+1} \geq \frac{B(I-i)}{\epsilon \lambda_{l_j}}, \qquad l_j \leq i < l_{j+1},$$

imply $\widetilde{u}_i - \widetilde{u}_{i+1} = \beta_{l_j} = B(I - l_j)/\epsilon\lambda_{l_j}$, the rent associated with the lowest number of defaults in the tranche. Consequently, the sponsor's payoff in state N is

$$\widetilde{u}(N) = \beta_{l_j}(l_{j+1} - N) + \sum_{k=j+1}^{n} \beta_{l_k}(l_{k+1} - l_k)$$

where j is the index of the tranche containing N. The continuation payoff at $t = 0$ is $\widetilde{u}_0 = \sum_{k=0}^{n} \beta_{l_k}(l_{k+1} - l_k)$.

To replicate $\widetilde{u}(N)$ consider all credit-linked notes[9] associated with CDO tranches $[l_k, l_{k+1})$ in L, yielding protection

$$P^k(N) = [N - l_k]^+ - [N - l_{k+1}]^+ . \tag{4}$$

[9] A credit-linked note refers to a funded structure as opposed to a "CDS style" deal where the protection is paid as and when a credit event occurs; cf. Chaplin (2005).

The protection embedded in a portfolio of CLNs $[l_k, l_{k+1})$, each with size β_{l_k}, is

$$\begin{aligned} P(N) &= \sum_{k=0}^{n} \beta_{l_k} P^k(N) = \sum_{k=0}^{j-1} \beta_{l_k}(l_{k+1} - l_k) + \beta_{l_j}(N - l_j) & (5)\\ &= \widetilde{u}_0 - \widetilde{u}(N), & (6) \end{aligned}$$

where j is defined as before. Thus $u(N) \leq \widetilde{u}(N) = \widetilde{u}_0 - P(N)$. We have shown the following.

Proposition 3 *With tranche-dependent default intensities, the sponsor's compensation $u(N)$ is more than hedged by the residual payoff of a weighted portfolio of CLN tranches $[l_k, l_{k+1})$ in L, each with nominal size*

$$\beta_{l_k} = \frac{B\,(I - l_k)}{\epsilon \lambda_{l_k}}.$$

The intuition is as follows. At date 0, the sponsor guarantees the deal for the trust by pledging $\widetilde{u}_0$ out of the proceeds of the sale. The funds are invested in a suitably weighted portfolio of CLNs underwritten by the trust to cover portfolio losses from 0 to $i^* + 1$ as above. The cost for the sponsor is the notional exposure of the CLNs, $\sum_k \beta_{l_k}(l_{k+1} - l_k) = \widetilde{u}_0$. At date 1, the protection embedded in the CLNs is $P(N)$. The sponsor quits the deal with the residual value of the CLNs, $\widetilde{u}_0 - P(N) = \widetilde{u}(N)$, more than the optimal payoff desired.

Next we deal with the payment of fees. The servicing fee,

$$\delta^1(N_t) = \lambda(N_t)\left(\widetilde{u}(N_t) - \widetilde{u}(N_t + 1)\right) = \sum_{k=0}^{n} \lambda_{l_k} \beta_{l_k} 1_{\{l_k \leq N_t < l_{k+1}\}},$$

is constant in each tranche and equal to $B(I - l_k)/\epsilon$. It is based on the maximum number of loans in the tranche rather than the nominal outstanding. The rent-preserving fee $\delta^2(N_t) = r\widetilde{u}(N_t)$ can accrue on the sponsor's account as follows.

Proposition 4 *With tranche-dependent default intensities, the rent-preserving fee is hedged by apportioning r to the sponsor and leaving $\Sigma_k - r$ with the trust out of all premiums Σ_k from the weighted CLN portfolio.*

The sponsor has to pledge $\widetilde{u}_0$ as a hedge against future liquidity risk. It is not entitled to the premium flows generated by the credit enhancement it provides, lest it is considered as a simple arbitrageur operating in the credit derivative market. Thus, all CLN premiums should stay with the trust. However, the sponsor may be more impatient than investors. Earmarking r uniformly out of all CLN tranches allows the trust to cater for the sponsor's impatience with the same instruments as those used for the synthetic compensation. The sponsor gets $r\widetilde{u}(N)$, the required rent-preserving fee. The intuition behind the outcome is that the premium "waterfall" gives priority to the payment of tranches more senior than the equity tranche. Payments to the latter tranches — in addition to the servicing fee already in place — ensure that the sponsor's informational rent remains high in the beginning and eventually tapers off as losses unfold. To induce the servicer to work in investors' best interest and conduct due diligence on borrowers, the sponsor must not only retain the first-loss piece, but extend protection on all tranches up to the senior piece, which goes all the way to i^* in our setup.

Last, we must determine which gain on sale S ensures that the trust beaks even. Since the sponsor prefunds the protection $\widetilde{u}_0$ out of the proceeds of the sale, the trust's initial cost is $I + S = \widetilde{u}_0 + \widetilde{v}_0$, where $\widetilde{v}_0$ is the liquidity generated by the securitization. There are deadweight costs from the implementation, because the sponsor receives more than the optimal fees and final payoff. With $\widetilde{v}_0 \leq v_0$, the share of the pool left for the sponsor's own fund, $K = I - \widetilde{v}_0$, is more than the capital required and the gain on sale S less than the social return under the optimal plan. We can ascertain the difference between the optimal and the super-replication strategy.

Proposition 5 *With tranche-dependent default intensities, the sponsor's and trust's continuation utilities associated with the super-replication strategy are respectively*

$$\begin{aligned} \widetilde{u}_0 &= \sum_{i=0}^{i^*} \beta_{l(i)} = \sum_{k=0}^{n} \beta_{l_k}(l_{k+1} - l_k) \\ \widetilde{v}_0 &= \sum_{i=0}^{i^*} (I-i)Rq_i - r\frac{B(I-l(i))}{\epsilon\lambda_{l(i)}} E\tau_i \wedge 1 - \widetilde{u}_0, \end{aligned}$$

where $l(i)$ is the index of the tranche containing i.

Compare with Proposition 2. The convex schedule $u(N)$ is approximated by a piecewise linear $\widetilde{u}$ with the same slopes at attachment points. The distortion implied by constant risk sharing in tranches is met by a higher credit enhancement $\widetilde{u}_0$ by the sponsor and a lower liquidity funding $\widetilde{v}_0$ by investors than would be available under the optimal plan. On the other hand, the super-replicating strategy can be implemented with standard credit derivatives instruments, is incentive-compatible, and ensures that the trust breaks even. The implementation could benefit from a deposit insurance fund to safeguard the trust against losses in its investment. The liability of the DIF at time 1 is the trust's shortfall in net income, if any, with respect to its credit derivative positions. (There is no income if $N_1 > i^*$, and both fees are at the top of the flow of funds.) The risk-related premium charged at date 0 could be fairly priced in the absence of asymmetric information between the trust and the fund.

4 Policy implications

The cost of mortgage debt has increased dramatically in recent months. Ouside investors and overseas buyers have now backed away following concerns about the US housing market and uncertainty about the involvement of the US government in the backing 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.[10] 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),

[10] 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 diversions 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 forward 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.[11] 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, whereas according to the model it cannot be securitized at all and may be held neither by remote investors nor by sponsoring banks. The fact that basis correlation can be found to be as high as one in the current environment seems indicative of faulty system design.

[11] 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.

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). Moreover, with forward monitoring, banks no longer have an incentive to enforce contracts after the loans are sold.

A related point 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 utility is still high but the fees 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 third 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[12] 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 results of the paper suggest that capital requirements alone may fail to 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 which sponsors are entitled, 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, in credit linked note rather than CDS style form, 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 instead of being paid to the protection seller. Only the fraction corresponding to rent-preserving fees is returned to sponsors with higher discount rates than investors. The remainder is withheld by the trust as the opportunity cost of the protection they receive for the systemic risk they create.

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 aggregate 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 resulting from their limited long term risk share. When losses begin unfolding capital is automatically supplied by originating banks and the tax is high. Only in case of a systemic crisis capital is overwhelmingly supplied by the trust and the liquidity tax eventually eschewed.

[12] Although our model does not consider loan cash flows before final redemption, the cashflow "waterfall" implied by actual CDOs usually allocates loan income from the reserve account 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.

5 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.[13] 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 does not fit in well however with current recommendations that the bankrupty code should be amended to allow for regulatory intervention ahead of bank insolvency. It suggests on the contrary that market discipline might be better imposed by a well-designed credit enhancement scheme based on tranching than outsourced to regulatory supervision.

The model suffers from clear limitations. Its results rely on an MLRP assumption which, in spite of its intuitive appeal, may very well prove wrong in a Markovian contagion framework with time homogeneous default intensities. They also proceed under the assumption that the parties to the securitization process

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

have the ability to commit and neglect, in particular, the bank's incentive to engage in optimal CRT once funds have been raised. Further issues would be worth exploring formally. We wish to show later that, in the absence of commitment, the bank might not be efficient. This would give an alternative interpretation of the current crisis beyond the motto of "immoral" bank behavior. We would also like to see how the time inconsistency problem could be addressed through capital requirements on loans conditioned on the extent of retained risk. This is left for future research.

6 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.

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, 2006, " Agency and optimal investment dynamics," *The Review of Financial Studies*, 20, 151–89.

Demyanyk, Yuliya and Otto Van Hemert, 2008, "Understanding the subprime mortgage crisis." Available at SSRN: http://ssrn.com/abstract=1020396.

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.

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.

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.

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

Franke, Günter and Jan Krahnen, 2006, "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.

Gorton, Gary and Nicholas Souleles, 2005, "Special purpose vehicles and securitization," Federal Reserve Bank of Philadelphia, Working Paper No 05-21.

Higgins, Eric and Joseph Mason, 2004, "What is the value of recourse to asset-backed securities? A study of credit card bank ABS rescues," *Journal of Banking and Finance*, 28, 857–874.

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.

Li, David, 2000,"On default correlation: A copula function approach," *Journal of Fixed Income*, 9, 115–18.

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.

7 Proofs

Proof of Proposition 1. Substituting for $\dot{u}_i$ in the dynamic programming equation, we see that δ_i must be chosen so as to maximize

$$\frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial u}\left(ru_i + \lambda_i(u_i - u_{i+1})\right) + \lambda_i\left(v_{i+1}(t, u_{i+1}) - v(t, u_i)\right) - \delta_i\left(1 + \frac{\partial v_i}{\partial u_i}\right).$$

Note that $u_I(t) = 0$. Assume by induction that $u_{i+1}(t) = u_{i+1}$ is independent of time. We take cases.

A. $\partial v_i / \partial u > -1$. Then $\delta_i = 0$ leads to the partial differential equation

$$\frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial u}\left(ru + \lambda_i(u - u_{i+1})\right) - \lambda_i v_i = \lambda_{i+1} v_{i+1}(t, u_{i+1})$$

with limit condition

$$v_i(1, u) = (I - i)\, R - u.$$

The general solution is

$$v_i(t, u) = e^{\lambda_i t} f\left(e^{-(r+\lambda_i)t}\left(u - \frac{\lambda_i u_{i+1}}{r + \lambda_i}\right)\right) - \int_t^1 \lambda_i e^{\lambda_i(1-s)} v_{i+1}(s, u_{i+1})\, ds$$

where f is determined by the limit condition above. This yields

$$\begin{aligned} v_i(t, u) &= e^{-\lambda_i(1-t)}\left[(I - i)R - \frac{\lambda_i u_{i+1}}{r + \lambda_i} - e^{(r+\lambda_i)(1-t)}\left(u - \frac{\lambda_i u_{i+1}}{r + \lambda_i}\right)\right] \\ &\quad - \int_t^1 \lambda_i e^{\lambda_i(1-s)} v_{i+1}(s, u_{i+1})\, ds, \end{aligned}$$

but this implies $\partial v_i / \partial u = -e^{r(1-t)} \leq 1$ as long as $r \geq 0$, a contradiction.

B. $\delta_i > 0$. Then $\partial v_i / \partial u = -1$ so that investors' continuation utility is linear. Let $v_i(t, u) = w_i(t) - u$. Substituting for v_i in the dynamic programming equation and using the induction $v_{i+1}(t, u_{i+1}) =$

$w_{i+1}(t) - u_{i+1}$, we find that w_i and u are related by

$$\dot{w}_i(t) - \lambda_i \left(w_i(t) - w_{i+1}(t)\right) = ru.$$

Since t and u are independent variable this is possible only when the left and right-hand sides are constant. Thus $u_i(t) = u_i$ and v_i is maximized by choosing u_i as low as possible, yielding

$$u_i = u_{i+1} + \frac{B(I-i)}{\epsilon\lambda_i}.$$

If $u < u_i$ the incentive constraint is violated and the bank cannot operate in state $N_t = i$. If $u > u_i$, the bank gets an immediate payment of $u - u_i$ and a dividend flow of

$$\delta_i = ru_i + \lambda_i(u_i - u_{i+1}) = ru_i + B(I-i)/\epsilon.$$

The bank's continuation utility is maintained at the minimum level u_i until maturity as long as $N_t = i$. The joint continuation utility $w_i(t)$ in state $N_t = i$ is given by $\dot{w}_i(t) - \lambda_i \left(w_i(t) - w_{i+1}(t)\right) = ru_i$ with limit condition $w_i(1) = (I-i)R$.

■

Proof of Proposition 2. Let k be some cut-off rule. Integrating the recurrence equations for the bank's continuation utility gives

$$u_0 = \sum_{i \le k} \frac{B(I-i)}{\epsilon\lambda_i}.$$

Let u and w be the vectors associated with u_i and w_i, respectively, for $i = 0, \ldots, k$. We know from Proposition (1) that

$$\dot{w}(t) + \Lambda w(t) = ru,$$

where

$$\Lambda = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \dots & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ & & \ddots & -\lambda_{k-1} & \lambda_{k-1} \\ 0 & \dots & & 0 & -\lambda_k \end{pmatrix}.$$

Integration yields

$$w(t) = \exp\left(\Lambda(1-t)\right)\left(\mathcal{R} - r\Lambda^{-1}u\right) + r\Lambda^{-1}u,$$

where $\mathcal{R} = w(1)$ is the vector of terminal conditions $w_i = R(I-i)$. On the other hand the forward Kolmogorov equations give for the corresponding vector of state probabilities $\dot{q}'(t) = q'(t)\Lambda$, which implies $q'(t) = (1,0,\dots)\exp(\Lambda t)$. Since

$$\Lambda^{-1} = \begin{pmatrix} -\lambda_0^{-1} & \dots & -\lambda_i^{-1} & \dots & -\lambda_k^{-1} \\ 0 & \ddots & \vdots & & \vdots \\ \vdots & 0 & -\lambda_i^{-1} & & \\ & & & \ddots & \\ 0 & \dots & & 0 & -\lambda_k^{-1} \end{pmatrix}.$$

we get

$$\begin{aligned} w_0 &= (1,0,\dots)w(0) = q'(1)\left(\mathcal{R} - r\Lambda^{-1}u\right) + r(1,0,\dots)\Lambda^{-1}u \\ &= \sum_{i\le k} q_i\left[R(I-i) + r\sum_{j=i}^{k}\frac{u_j}{\lambda_j}\right] - r\sum_{i\le k}\frac{u_i}{\lambda_i} \\ &= \sum_{i\le k} q_i R(I-i) - r(1-Q_i)\frac{u_i}{\lambda_i} \end{aligned}$$

using Fubini with $Q_i = \sum_{j \leq i} q_j = P(N_1 \leq i)$. Substituting for u_i we find

$$\begin{aligned} w_0 &= \sum_{i \leq k} q_i R(I-i) - r \frac{1-Q_i}{\lambda_i} \sum_{j=i}^{k} \frac{B(I-j)}{\epsilon \lambda_j} \\ &= \sum_{i \leq k} (I-i) \left(q_i R - \frac{rB}{\epsilon \lambda_i} \sum_{j \leq i} \frac{1-Q_j}{\lambda_j} \right) \end{aligned}$$

using Fubini again.

Consider $Q_i(t) = \sum_{j \leq i} q_j(t) = P(N_t \leq i)$. From the forward Kolmogorov equations, we know that $\dot{Q}_i = -\lambda_i q_i$ so that integrating from 0 to 1

$$\frac{1-Q_i}{\lambda_i} = \int_0^1 q_i(t)\, dt.$$

We conclude that

$$\begin{aligned} w_0 &= \sum_{i \leq k} (I-i) \left(q_i R - \frac{rB}{\epsilon \lambda_i} \int_0^1 Q_i(t)\, dt \right) \\ &= \sum_{i \leq k} (I-i) \left(q_i R - \frac{rB}{\epsilon \lambda_i} E \int_0^1 1_{\{t \leq \tau_i\}}\, dt \right) \\ & \sum_{i \leq k} (I-i) \left(q_i R - \frac{rB}{\epsilon \lambda_i} E\tau_i \wedge 1 \right), \end{aligned}$$

implying that investors' continuation value at time 0 is

$$v_0 = \sum_{i \leq k} \frac{R(I-i)}{\lambda_i} \left(\lambda_i q_i - \frac{B}{R\epsilon}(1 + rE\tau_i \wedge 1) \right).$$

We now show that the sequence $\lambda_i q_i$ is first increasing and eventually decreasing. Under MLRP,

$$\frac{q_0'}{q_0} \leq \cdots \leq \frac{q_I'}{q_I},$$

where the prime denotes the probability in the absence of monitoring. The explicit formula for q_i with $i < I$

takes the form

$$q_i(t) = \sum_{j=0}^{i} a_{i,j} e^{-\lambda_j t}$$

for appropriately chosen coefficients $a_{i,j}$ that are homogeneous of degree zero in λ. Thus $q_i'(1) = q_i(1+\epsilon)$ for all $i \le I$. Since $q_i'/q_i = \exp \int_1^{1+\epsilon} (\dot{q}_i/q_i)\, dt$, MLRP holds true for any ϵ if and only if $\dot{q}_i/q_i \le \dot{q}_{i+1}/q_{i+1}$ for all i. But

$$\sum_{i=0}^{I} q_i \frac{\dot{q}_i}{q_i} = 0$$

implies that $\dot{q}_i/q_i$ is first negative and then positive. The forward Kolmogorov equations

$$\lambda_i q_i - \lambda_{i-1} q_{i-1} = -\dot{q}_i$$

imply in turn that the sequence $\lambda_i q_i$ is first increasing and then decreasing after some rank, as desired.

Returning to the definition of v_0, we see that the first term in the summation is positive by assumption, while the other terms come from the difference between two sequences, one eventually decreasing and the other increasing. Clearly, their contribution must be negative after some rank. The optimal cut-off rule i^* is the maximum i such that

$$\frac{B}{\epsilon \lambda_i}(1 + rE\tau_i \wedge 1) \le Rq_i.$$

■

Proof of Proposition 4. Consider the CDO tranche with attachment points l_k, l_{k+1} and premium Σ_k. The premium leg is

$$\begin{aligned} CDO_t^k &= \Sigma_k \left(\int_0^{\tau_{l_k - 1} \wedge t} (l_{k+1} - l_k)\, dt + \int_{\tau_{l_k - 1} \wedge t}^{\tau_{l_{k+1} - 1} \wedge t} (l_{k+1} - N_t)\, dt \right) \\ &= \frac{\Sigma_k}{\beta_{l_k}} \left(\int_0^{\tau_{l_k - 1} \wedge t} (\widetilde{u}_{l_k} - \widetilde{u}_{l_{k+1}})\, dt + \int_{\tau_{l_k - 1} \wedge t}^{\tau_{l_{k+1} - 1} \wedge t} (\widetilde{u}(N_t) - \widetilde{u}_{l_{k+1}})\, dt \right). \end{aligned}$$

Defining $\alpha_k = r\beta_{l_k}/\Sigma_k$ and summing over k we get

$$\begin{aligned}\sum_{k=0}^{n}\alpha_k CDO_t^k &= r\sum_{k=0}^{n}\left(\widetilde{u}_{l_k}\,\tau_{l_k-1}\wedge t-\widetilde{u}_{l_{k+1}}\,\tau_{l_{k+1}-1}\wedge t+\int_{\tau_{l-1}\wedge t}^{\tau_{l_{k+1}-1}\wedge t}\widetilde{u}(N_t)\,dt\right)\\ &= r\int_0^{\tau_{k^*}\wedge t}\widetilde{u}(N_t)\,dt = r\int_0^t\widetilde{u}(N_t)\,dt,\end{aligned}$$

since $\tau_{-1}=0$, $\widetilde{u}_{l_n}=\widetilde{u}_{i^*+1}=0$ and $\widetilde{u}$ vanishes after i^*, as desired. ■

Proof of Proposition 5. The trust has all control rights and gets $RE\,[I-N_1;N_1\le i^*]$ at maturity. In addition, it receives the premiums Σ_k of the CLNs, less the fee r charged by the sponsor to preserve its rent. If CDO_t^k denotes the cumulative premium flow associated with a single tranche $[l_k,l_{k+1})$, the absence of arbitrage between the protection and premium legs at $t=0$ implies that

$$E\sum_{k=0}^{n}\beta_{l_k}CDO_1^k = E\sum_{k=0}^{n}\beta_{l_k}P^k(N_1) = EP(N_1)=\widetilde{u}_0-E\widetilde{u}(N_1).$$

By Proposition 4, the premium remitted to the sponsor is

$$E\sum_{k=0}^{n}\beta_{l_k}\frac{r}{\Sigma_k}CDO_1^k = rE\int_0^1\widetilde{u}(N_t)\,dt.$$

Finally, the trust must compensate the sponsor with expected fee of $E\int_0^1\delta^1(N_t)\,dt$ for servicing. Gathering terms,

$$\begin{aligned}&E\left[R\,(I-N_1)\,1_{\{N_1\le i^*\}}+\sum_{k=0}^{n}\beta_{l_k}\,(1-r/\Sigma_k)\,CDO_1^k-\int_0^1\delta^1(N_t)\,dt\right]\\ = \;&\widetilde{u}_0+\sum_{i=0}^{i^*}R(I-i)q_i-E\left[\widetilde{u}(N_1)+\int_0^1\delta^1(N_t)\,dt+r\int_0^1\widetilde{u}(N_t)\,dt\right].\end{aligned}$$

From the sponsor's integrated promise-keeping constraint, we know that

$$\widetilde{u}_0=E\left[\int_0^1 e^{-rt}\left(\delta^1(N_t)+r\widetilde{u}(N_t)\right)\,dt+e^{-r}\widetilde{u}(N_1)\right],$$

implying in particular that $E\left[\widetilde{u}(N_1) + \int_0^1 \delta^1(N_t)\,dt\right] = \widetilde{u}_0$. Moreover

$$\begin{aligned} E\left[r\int_0^1 \widetilde{u}(N_t)\,dt\right] &= E\left[\int_0^1 \sum_{i=0}^{i^*} r\widetilde{u}_i 1_{\{N_t=i\}}\,dt\right] \\ &= E\left[\int_0^1 \sum_{i=0}^{i^*} r(\widetilde{u}_i - \widetilde{u}_{i+1})1_{\{N_t\le i\}}\,dt\right] \\ &= \sum_{i=0}^{i^*} r(\widetilde{u}_i - \widetilde{u}_{i+1})E\tau_i \wedge 1. \end{aligned}$$

The net expected value of the trust's investment is

$$\sum_{i=0}^{i^*}(I-i)Rq_i - r(\widetilde{u}_i - \widetilde{u}_{i+1})E\tau_i \wedge 1 = \widetilde{u}_0 + \widetilde{v}_0,$$

with the definitions given in the Proposition. ■