Loan servicers’ incentives and optimal CDOs

Henri Pagès*
Banque de France, Research Department
November 23, 2009

Abstract

The paper examines a delegated monitoring problem between investors and servicers managing a pool of correlated loans subject to Markovian “contagion.” Moral hazard induces a foreclosure bias in the decision of loan servicers unless they are compensated with the right incentive-compatible contract. The asset pool is liquidated when losses exceed a state-contingent cut-off rule. Servicers bear a relatively high share of the risk initially, as they should have high-powered incentives to renegotiate, but their long term financial stake tapers off as losses unfold. Liquidity regulation based on tranching 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 rewards servicers with servicing and rent-preserving fees for outcomes that signal diligent servicing, i.e., 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, Securitization, Default Risk, Servicer, 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, Diana Bonfim, Gilles Chemla, Gabriella Chiesa, Esa Jokivuolle, Antoine Martin, Jean Tirole as well as participants in seminars at the Bank of France, the Bank of Finland, IDEI and the CREDIT 2009 conference in Venice 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 disclaimer applies.
1 Introduction

Servicers develop specialized collection skills on behalf of investors in exchange for investors’ ability to fund their activities. Failure to commit adequate loss-mitigation techniques results in high foreclosure rates, which can ultimately jeopardize financial stability. This buy-side agency problem can be felt acutely in subprime mortgage markets if foreclosure rates have negative externalities on housing prices which are aggravated when fundamentals go wrong. The focus of the paper is on how contractual arrangements between servicers and investors interact with strategic behavior in determining foreclosure rates when contagion is present.

One novel feature of the approach is to show that the delegated monitoring problem between risk-neutral investors and servicers 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, there would barely be any servicing or management function to describe. However, the complex nature of ABS transactions introduces 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 servicers can be as important to the performance of the pool as its initial composition and characteristics. The recent crisis in the US subprime mortgage market has shifted attention to the crucial foreclosure decision of servicers. As investors’ agents, servicers may not struggle hard to renegotiate delinquent loans once a major stake in them has been sold. Piskorski et al (2009) show that, conditional on being seriously delinquent, loans are more likely to be foreclosed when securitized rather than held by the bank. The incentives for quick foreclosure may come from various frictions, such as the low flexibility allowed for by the original Pooling and Servicing Agreement, the presence of substantial un-reimbursable expenses incurred during renegotiation or, more generally, the failure to internalize the costs and benefits of the decision to foreclose.
Given the reduction in accountability implied by securitization, one should not trust that servicers act necessarily on investors’ best interest. A mechanism design analysis of optimal securitization can be helpful in examining moral hazard in the time dimension and seeing how frictions that preclude loan renegotiation are minimized. 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.

To shed light on the dynamic delegated monitoring problem, we start with a stylized model where servicers 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, servicers can make a costly effort at any point in time to service diligently, in which case the loss intensity of the pool improves at that time. Given competitive investors, the goal is to elicit which high-powered compensation maximizes servicers’ payoff subject to a zero-profit condition for investors and an incentive compatibility condition for servicers. The optimal contract allows for some risk sharing while maintaining diligent servicing throughout.

The paper considers a single pool and cannot account for systemic interdependence in the financial sector. For this reason servicers’ exposure to systemic risk is taken as exogenous and handled with a Markovian model of “contagion.” A convenient reduced-form approach is to assume that the underlying loss intensity is governed by the arrival of default. Laurent, Cousin and Fermanian (2008) show that a loss distribution calibrated either on a one-factor Gaussian copula with given correlation or on a typical base correlation curve for the iTraxx generates intensities which increase with the number of defaults. This suggests that a Markovian model, where the intensity dynamics are taken as inputs, is well suited to capture the dependence structure observed during contagion episodes. Low probability high impact events are 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 losses start being lumped together.
To simplify the exposition we consider a static portfolio of identical, infinitely-lived defaultable loans yielding constant cash flows per unit time. The optimal risk prevention policy relies on two instruments: positive payments to servicers 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 servicers. 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 servicers work in their best interest, investors resort to the carrot-and-stick approach. They are rewarded when the track record is on target. Two kind of fees are charged in the “bliss” state. One is the direct servicing fee which is a flat percentage of the outstanding loan balance. The other is a rent-preserving fee for impatience which depends on servicers’ discount rate. When the track record deteriorates, however, payments are suspended. Servicers take stick from investors through a reduction in their continuation utility when a loss occurs. The magnitude of the “penalty” is pinned down by the incentive compatibility constraint. In the beginning underlying risk is low and it is difficult to disentangle diligent servicers from careless ones. They need high-powered incentives and bear the brunt of initial losses. In the end underlying risk is high and the likelihood of subsequent defaults makes servicers eager to apply their collecting skills. They are 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, servicers’ risk share tapers off as losses unfold, until the pool is exhausted.

In the model, the optimal retention rate requirement is linked to the incidence of systemic risk. In contrast with the recent amendment of the capital requirement directives, which asks sponsors to keep a minimum “one-size-fits-all” 5% retention rate in a securitization sold to investors that include a credit institution — in practice a minimum “vertical strip” of each tranche — the optimal plan has declining risk shares. They are consistent with the interpretation put on the information intensity of securities by Gorton and Pennacchi (1990) or Dang et al (2009). Initially, sponsors should hold relatively many information-intensive
junior tranches, as affiliated servicers need incentives to acquire information about borrowers. Eventually, once doubts about the quality of borrowers are paramount, sponsors should hold relatively few information-insensitive senior tranches, as they must not take advantage of the privileged information affiliated servicers have.

Next consider the stochastic liquidation policy. Penalties meted out after losses define servicers’ reservation utility. When continuation utility comes close to reservation level, the threat of reductions has no real bite because servicers are protected by limited liability. To cope with the situation, investors allow for random liquidation\(^1\) of the pool upon default, with a probability of survival reflecting servicers’ current performance. The threat of liquidation impels them to keep servicing diligently even if their 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 high-performing servicers are allowed to make without fearing liquidation. Tuning the cut-off rule is as effective an instrument to discipline servicers as the punishment itself.

To understand the mechanics of the cut-off rule, recall that once on probation servicers are driven by the prospect of future payments. As long as there are no losses, payments should be resumed soon and the new performance target adjusted to assuage servicers’ longing for fees. In normal circumstances — assuming individual risk is not noticeably affected — it is not sensible to keep servicers waiting with the promise of larger payments since underlying conditions have not changed. The reason for actually reducing payments is twofold. First, compensation should not improve in size-adjusted terms and the pool has decayed by one unit. Second, risk shifting incentives should be held in check and foreclosures 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, high-performing servicers know that they will lose after a loss even if they do not have to wait long and remain diligent.

\(^1\)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.
But there is a twist. Individual risk may surge in rare circumstances. It is then that servicers’ special skills at collecting payments are most valuable. The aggregate loss intensity soars despite the reduction in pool size and dwarfs servicers’ 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, high-performing servicers know that they will have to wait a long time before payments are resumed if they come to operate under turbulence and remain diligent.

Interestingly, the optimal policy can be implemented with tranching when changes in individual risk are lumpy. To this extent the complex institutions involved in structured finance can sow the seeds of their salvation. More specifically, the sponsor sells the pool to a trust and guarantees the deal by returning the capital required and gain on sale to a reserve account managed by the trust. It then hires the servicers of the pool. The internal credit enhancement is used as cash collateral to reimburse the trust for losses according to servicers’ optimal risk shares, following a portfolio of suitably weighted collateralized debt obligations (CDOs). Fees accrue on the reserve account to increase credit support and are remitted to servicers when the balance is on target. In contrast, CDO premium spreads are retained by the trust as a liquidity tax for the systemic risk assumed. Movements in the reserve account faithfully mirrors servicers’ performance and are used by the trust to trigger stochastic liquidation. The result shows that the optimal tradeoff between efficient risk sharing and diligent servicing is consistent with separating different functions, with servicers affiliated with the sponsor on the one hand, and parties related to the securitization on the other.

The intuition behind this result is that CDOs are an ideal way to make non-standardized or non-traded loans more transparent. The weights of the different tranches reflect the penalty servicers should face when the corresponding tranches become active. Losses are revealed to dispersed investors as they occur and act as a trigger mechanism to make margin calls. The main difference between margin calls and our mechanism is liquidity management. In a margin account, the broker asks the investor to post new collateral as prices drop. In a reserve account, the cost of the protection is posted in advance and covered by withdrawals as
losses unfold. The margin calls can be determined from the pricing of CDO tranches. Since each CDO is a “bet” on the portfolio making losses in a specific tranche, their prices reflect the underlying loss intensities as perceived by the market at the time of securitization.

In essence, liquidity management follows servicers’ performance record and removes the uncertainty about the underlying quality of servicing. The reserve account balance is maintained between size-contingent minima and maxima and meant to be used up in case of need. Drawing down the liquidity position does not leave the buffer exposed to the “repeated” liquidity shock conundrum (Goodhart, 2008). Should a jolt of bad news deplete the reserve account below the minimum required, the pool goes into stochastic liquidation with two options: replenishment by the trust, or takeover by the regulator. The current balance yields the market value of performance, thus obviating the need to rely on a proprietary quantitative risk model or some expert judgement exercised by bank supervisors. The pool does not have to be adjusted to the requirements of capital regulation. Instead, the capital buffer is automatically downsized as new information is revealed. Hence, there is no risk of a countercyclical effect of capital requirements on insolvency risk.

Such liquidity management is broadly consistent with the view that information acquisition plays a role in equilibriums with adverse selection when there is uncertainty about the motives for selling assets (Malherbe, 2009). A high liquidity position is justified when performance is good, as investors are not wary of the agent’s overhoarding liquidity through the securitization of low quality assets. Conversely, a tight liquidity position is justified when performance is poor, as the threat of stochastic liquidation assuages fears about the quality of servicing and dispenses with the need for costly liquidity hoarding.

The paper is organized as follows. Section 2 offers a brief review of the related literature. In Section 3, we present the model and characterize the optimal contract under the incentive compatibility and limited liability constraints. Based on this analysis, Section 4 adopts a backward recursive approach to construct the solution of a system of optimal control problems and derive the dynamics of pool size. Section 5 offers some tentative policy implications before the conclusion in Section 6. An Appendix contains details about the analytics of pool size and the proof of Proposition 3.

2Relatedly, Hart and Zingales (2008) suggest that the price of CDS should be used to ensure that banks maintain an adequate capital buffer.
2 Related literature

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 continuous 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) discuss the array of frictions that arise in the atomized setting of securitization and show that the servicer’s role is not confined to the collection and remittance of loan payments. It carries important responsibilities, like maintenance of property, advance payments or escrow administration when the loan starts being delinquent and foreclosure management once deemed uncollectible. These activities have consequences for the performance of loans, with an impact of plus or minus 10 percent on loss according to a Moody’s estimate. Piskorski at al (2009) show empirically that securitization induces a foreclosure bias in private subprime mortgages, with delinquent loans having a foreclosure rate between 3.8 and 7 percent lower in absolute terms (18 to 32 percent in relative terms) when held by the bank rather than securitized. Their finding is robust to the inclusion of unobservable characteristics that lenders may obtain during origination or through
subsequent monitoring. This contrasts somewhat Adelino et al (2009) who find on a smaller sample of the same dataset that servicers do not use direct modifications of contractual terms as a frequent renegotiation tool. 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. Pennington-Cross and Ho (2006) examine the heterogeneity of servicers in securitized subprime mortgages and estimate very large increases or decreases in the probability of a loan going to default or prepayment relative to the reference group.

Several papers examine the implications of credit risk transfer (CRT) for banks’ incentives to monitor that recent empirical studies document. 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. 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.

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 al-

---

3See, for example, Berndt and Gupta (2008), Drucker and Puri (2007), Keys et al. (2008).
locating information-insensitive derivative securities to remote 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 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.

3 The model

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. 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 continuous servicing has two consequences. First, the distinction between the exogenous base quality of the loans and the endogenous default probability that obtains after the servicing decision has taken place arises at each point in time. Second, the cost of servicing depends on how defaults propagate in the portfolio. We rely on a homogeneous “contagion” Markov model where the loss intensity of the $n$th-to-default loan depends on the size of the pool. We show that, if investors can commit, they will ensure that
servicers are diligent by subordinating fees to performance and winding down the pool when losses exceed a state-dependent threshold.

3.1 Continuous servicing

Consider a pool of securitized loans held by a trust on behalf of investors and administered by servicers. The former have unlimited liability and supply liquidity competitively, as long as they cover the costs. The latter are the sole entity which borrowers interact with. They are affiliated with a sponsor which has funds to invest. There is universal risk neutrality and servicers are more impatient than investors. Investors’ objective is to find a contract that maximizes their expected profit subject to servicers collecting diligently. Knowing this, the sponsor will offer investors a contract that fetches as much as possible and lets investors break even.

The pool consists of $I$ unit loans, the default risk of which has some systematic component. All loans are ex ante identical and yield $\mu$ per unit time. The pool 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 not managed by servicers. This is meant to capture the idea that pool illiquidity stems in large part from the servicer-borrower relationship, implying that the ability to collect loans rests squarely on their unique skills at working with borrowers behind on payments and extracting more concessions from them.

Let $i = I - N_t$ be the size of the pool\(^4\) 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 default and liquidation. 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 $N_j^i$ have default intensities depending on the size of the pool and on the quality of servicing. If servicers are diligent, default intensities are $\alpha^i(t) = \alpha_i$ for the $i$ loans outstanding and zero otherwise. Thus, as long as the pool is spared from liquidation risk, the aggregate loss intensity is $\lambda_i = \sum_j \alpha^j(t) = i\alpha_i$.

\(^4\)To avoid cumbersome notation, the time index of portfolio size $i$ is systematically suppressed.
Servicing 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 servicers choose to shirk \((e_t = 0)\), they enjoy 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 \((e_t = 1)\), uniformly in \(i\). One interpretation of \(\epsilon\lambda_i\) is the foreclosure bias obtained when servicers apply their best loss-mitigation techniques to delinquent loans.

A contract specifies the amount \(\delta_t\) to be paid to servicers and the time \(\tau\) 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 \(\theta\), so the pre-liquidation intensity associated with the indicator \(M_t = 1_{\{t \geq \tau\}}\) is \(\lambda_i(1 + (1 - e_t)\epsilon)(1 - \theta)\). The sequence of events is as follows.

The size inherited from the past is \(i\). Servicers receive payment \(\delta_t\,dt\) and make effort decision \(e_t\) for \((t, t + dt)\). With probability \(\lambda_i(1 + (1 - e_t)\epsilon)\,dt\) there is a loss and the size becomes \(i - 1\). Then the pool is liquidated with probability \(1 - \theta\). Otherwise servicers keep administering the pool, with initial size or one less unit.

### 3.2 Incentive compatibility and limited liability

Let \(r\) be servicers’ 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 servicers’ lifetime utility at \(t\) as the conditional expected discounted revenue of their activities

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

given a contract \((\delta, \tau)\) and an effort process \(e\). Related to lifetime utility is continuation utility defined as

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

Servicers participate only if their continuation gains, plus any monetary and private dividends, match their impatience. Since

\[
U_t = \int_0^\tau e^{-rs} \left( \delta_s + (1 - e_s) B(I - N_s) \right) \, ds + e^{-rt} u_t
\]

12
is a martingale, the integral representation theorem for point processes (Brémaud, 1981) implies that there are predictable processes\(^5\) \(h^1\) and \(h^2\) such that the continuation utility satisfies the promise-keeping equation

\[
    du_t + (\delta_t + (1 - e_t) B(I - N_t)) dt = ru_t dt - h^1_t \left( \sum_j dN_j^t - \alpha^t_j (1 + (1 - e_t) \epsilon) dt \right) - h^2_t (dM_t - (1 - \theta) \lambda_i (1 + (1 - e_t) \epsilon) dt)
\]

(2)

until liquidation. The 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 \((\delta, \tau)\), choosing \(e_t = 1\) at each point in time is incentive compatible if and only if

\[
    h^1_t + (1 - \theta) h^2_t \geq b_i = \frac{B}{\epsilon \alpha^t_i},
\]

(3)

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

Heuristically, if servicers plan to follow the optimal strategy \(e = 1\) starting from \(t\), they should have no incentive to deviate before \(t\). From (1), their continuation utility \(u_t\) is determined by the history of defaults and the contract \((\delta, \tau)\) after time \(t\), not by effort before time \(t\). Given \(u_t\), they will not deviate between \(t - dt\) and \(t\) if the real change in continuous utility \(du_t - ru_t dt\) is lower under diligent servicing. This yields the incentive compatibility constraint (3). The left-hand side is the “penalty,” the predictable loss in utility brought about by default or liquidation risk. The right-hand side is the minimum rent consistent with diligent servicing in state \(i\), the usual non-pledgeable income reflecting the attractiveness of private benefits when the agent is shirking. It defines reservation utility, since below \(b_i\) utility would be less than the penalty imposed after a loss.

A high sensitivity to losses requires that servicers be compensated with high utility in the beginning. This reduces investors’ value. Hence, the incentive compatibility binds under the optimal plan. Because

\(^5\)Since outstanding loans are indistinguishable, we assume w.l.o.g. that \(h^1\) is a scalar process.
liquidation is inefficient and should be avoided to the extent possible, there are two regimes for servicers. Either \( u \geq b_i + b_{i-1} \) and there is no need to liquidate the pool (\( \theta = 1 \)). The loss in utility is \( h_i^1 = b_i \) 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 liquidated, the promise-keeping constraint yields \( u = h_i^1 + h_i^2 \). The incentive compatibility constraint in turn determines \( \theta = (u - b_i) / (u - h_i) \). But limited liability has \( u - h_i \geq b_{i-1} \) when the pool is spared, so \( \theta \) is maximized when \( h_i^1 = u - b_{i-1} \) and \( h_i^2 = b_{i-1} \). The optimal survival probability, \( \theta = (u - b_i) / b_{i-1} \), reflects servicers’ position in the interval \( [b_i, b_i + b_{i-1}] \).

If a default occurs, utility is first reduced to \( u - h_i = b_{i-1} \), servicers’ reservation utility in state \( i - 1 \). Then a loaded coin (probability \( \theta \)) is thrown. Heads servicers remain in charge and their utility starts growing. Tails the pool is liquidated and \( b_{i-1} - h_i = 0 \).

### 3.3 Optimal contracting

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

\[
\ddot{u}(t) + \delta_t = ru(t) + \lambda_i b_i \land (u(t) - b_{i-1}) + \lambda_i (1 - \theta) b_{i-1}
\]

\[
= ru(t) + \lambda_i b_i
\]

between two successive losses. Servicers charge two kinds of fees to investors. One shields them against the incidence of losses for which they are not accountable under diligent servicing. The servicing fee \( \lambda_i b_i = iB/\epsilon \) is a flat percentage of the outstanding pool. The other maintains the real value of their continuation utility and is tuned to the rate of impatience.

In this time-homogeneous setup, as usual in models of dynamic moral hazard, the current size of the pool \( i \) and servicers’ current utility \( u \) are sufficient statistics for the optimal contract. Investors’ continuation utility, \( v_i(u) \), satisfies the following system of Hamilton Jacobi Bellman equations which can be solved recursively
\[
\max_{\delta_t(\cdot)} \{ (ru + \lambda_i b_i - \delta_t) \dot{v}_i(u) + i\mu - \delta_t - \lambda_i \theta (v_i(u) - v_{i-1}((u - b_i) \lor b_{i-1}) - \lambda_i(1 - \theta)v_i(u) \} = 0,
\]
where \( \theta = [(u - b_i)/b_{i-1}] \land 1 \) is the optimal probability of liquidation given default and \( v_0(u) = 0 \). 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 servicers. The last two correspond to the loss of utility incurred depending on whether servicers keep 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

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

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 \( \delta_t (\dot{v}_i(u) + 1) = 0 \) helps explain why. When \( u \) is above target, social surplus \( u + v_i(u) \) is maximized and \( \dot{v}_i(u) = -1 \). Investors prevent \( u \) from rising above target by paying fees to servicers. Below target \( \dot{v}_i(u) > -1 \) 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 \( i \), servicers fear liquidation risk after a loss.

### 4 Pool size dynamics

With constant returns to scale, servicers’ reservation utility, \( b_i = B/(\epsilon \alpha_i) \), does not change as long as \( \alpha_i \) remains constant. The size-adjusted rent \( B/(\epsilon \lambda_i) \) edges up as the foreclosure bias declines with the number of loans outstanding and it becomes increasingly difficult to disentangle diligent servicers from careless ones. In contrast, when bouts of contagion trigger a sharp rise in underlying default risk, the foreclosure bias rises abruptly. They have less leeway to shirk. We are interested in the implications that such changes have for the design of the optimal contract.
Markovian contagion between defaults is introduced by assuming that the sequence $\alpha_i$ 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 trust’s ability to diversify credit risk and makes the last few loans comparable to “economic catastrophe bonds” (Coval et al, 1999), low in risk unconditionally but likely to be wiped out if the risk materializes.

We make the following assumptions.

**Assumption 1** The sequence $\alpha_i$ is decreasing: $\alpha_{i-1} \geq \alpha_i$.

As the size of the portfolio $i$ edges down over time, the underlying individual intensities $\alpha_i$ keep increasing or stay constant. Recall that aggregate intensity is $\lambda_i = i\alpha_i$. Since $\lambda_{i-1}/\lambda_i \geq (i-1)/i$, aggregate default intensity jumps with underlying individual risk and cannot decrease by more than $1/i$ if the latter is constant.

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

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

### 4.1 Single loan: Constant utility

Investors set servicers’ continuation utility at its minimum level $b_1$. This implies a continuous payment of $\delta_1 = b_1 (r + \lambda_1)$. In this degenerate special case the HJB equation returns

$$v_1 = \mu - \delta_1 \lambda_1,$$

the present value of $\mu - \delta_1$ until extinction at time $\tau$ since $E[\tau - t | F_t] = 1/\lambda_1$. Optimal policy is captured by the value function $v_1(u) = v_1 - (u - b_1)$ for $u \geq b_1$. When $u > b_1$, an immediate payment of $u - b_1$ is made to have servicers fall back on their reservation utility $b_1$. However, $u > b_1$ is never reached under the optimal plan.
4.2 Two loans: Stochastic liquidation

It no longer pays to limit servicers to their reservation utility. We know from the incentive compatibility constraint that when \( u \) belongs to \([b_2, b_2 + b_1]\), there is stochastic liquidation upon default with rate of survival \( \theta = (u - b_2)/b_1 \). Hence, servicers’ utility can be written as \( u = b_2 + b_1 \theta \) where \( 1 - \theta \) is the probability that the pool is liquidated if a default occurs. In the absence of payments, servicers’ continuation utility grows as

\[
\dot{u}(t) = ru(t) + \lambda_2 b_2 = ru(t) + 2B/\epsilon
\]

until \( \gamma_2 = b_2 + b_1 \) is reached. It is neither optimal to prevent \( u \) from increasing before \( \theta \) is equal to one, nor to keep it increasing beyond \( \gamma_2 \). The stochastic liquidation interval is exactly \([b_2, \gamma_2]\).

On \([b_2, \gamma_2]\) investors’ continuation utility \( v_2(u) \) satisfies the HJB equation

\[
(ru + \lambda_2 b_2) \dot{v}_2(u) + 2\mu - \lambda_2 (v_2 - \theta \pi_1) = 0,
\]

the solution of which\(^6\) can be written as \( v_2(u) = b_1 w_2^1 (\theta) \) for some normalized function \( w_2^1 \). The slope at \( \theta = 1 \) is given by the boundary condition \( w_2^1(1) = -1 \), ensuring that servicers get paid only when \( u = \gamma_2 \). Should default of the penultimate loan occur in that state, their utility jumps into the single loan regime \( u = b_1 \).

Investor’s continuation utility is concave on \([b_2, \gamma_2]\). This property, as in all solutions of higher size, reflects the inefficiency arising from stochastic liquidation. The principal’s value react all the more strongly to performance as liquidation is likely and the highest inefficiency arises when servicers are constrained at their reservation level. We assume that a higher performance of servicers originally raises investors’ continuation utility. A technical condition given in Appendix ensures that this is indeed the case irrespective of the size.

\(^6\) Details of this derivation, as of those in the subsections below, are given in the Appendix.
4.3 Three loans: One exemption from liquidation under probation

As above, stochastic liquidation arises in the first interval $[b_3, b_3 + b_2]$. Investors’ continuation utility can be written as $v_3(u) = b_2 u_3^2(\theta)$ where the survival probability $\theta = (u - b_3) / b_2$ is the position of $u$ in that interval. However, it is no longer optimal to prevent $u$ from exceeding the stochastic liquidation interval. Beyond $b_3 + b_2$ servicers must be let out on probation for some time. Let $\theta = (u_3 - b_3 - b_2) / b_1$ be servicers’ position in the second interval $[b_3 + b_2, b_3 + b_2 + b_1]$. One can show that investors’ utility is now given by $v_3(u) = b_1 u_2^3(\theta)$, where the upper boundary of probation $\overline{\theta}_3$ solves the two conditions

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

The first states that it is no longer cheaper to compensate servicers using future rewards rather than an immediate transfer. The second is a “smooth pasting” condition ensuring that $\overline{\theta}_3$ is optimal: if $w_3^2$ were strictly concave at $\overline{\theta}_3$, more surplus could be obtained by marginally raising the threshold beyond that level.

One finds after some substitutions that the critical level $\overline{\theta}_3$ lies strictly between 0 and 1 and is such that

\begin{equation}
1 + \dot{v}_2(b_2 + \overline{\theta}_3 b_1) = \frac{r}{\lambda_3}.
\end{equation}

The penalty is $b_3$. If a loss occurs during probation, servicers see their continuation utility drop to $u - b_3 = b_2 + \theta b_1$, so $\theta \in [0, \overline{\theta}_3]$ can be interpreted as the probability of survival given default that servicers face when their utility jumps into the two-loan stochastic liquidation interval following a loss. A useful characterization for the sequel is to notice that servicers are exempted from liquidation once during probation, since two consecutive defaults effectively end the game with strictly positive probability. In the absence of default $u$ keeps increasing until the social value of performance after the penalty is imposed, $1 + \dot{v}_2(u - b_3)$, equates its current relative cost $r / \lambda_3$ at $\theta = \overline{\theta}_3$. When $u$ reaches the target $\gamma_3 = b_3 + b_2 + \overline{\theta}_3 b_1$, the slope of investors’ continuation utility is $-1$ and servicers get their share in the form of fees.
4.4 Four loans: Up to two exemptions from liquidation under probation

With penalty $b_4$ stochastic liquidation arises in $[b_4, b_4 + b_3]$. The size of probation depends on a comparison between the current relative cost of performance, $r/\lambda_4$, and the social value of performance after the penalty is imposed, $1 + \dot{v}_3(u - b_4)$. We take cases.

4.4.1 Low aggregate loss intensity

Suppose first that aggregate loss intensity $\lambda_4$ is so low that $1 + \dot{v}_3(b_3 + b_2) \leq r/\lambda_4$. Since $v_3$ is strictly concave there exists $\overline{\theta}_4$ between 0 and 1 such that

$$1 + \dot{v}_3(b_3 + \overline{\theta}_4 b_2) = \frac{r}{\lambda_4}.$$ 

Then probation consists of a single interval $[b_4 + b_3, \gamma_4]$ with target defined as $\gamma_4 = b_4 + b_3 + \overline{\theta}_4 b_2$. If default occurs during probation, servicers’ utility drops in the stochastic liquidation interval of $i = 3$ with continuation utility $b_3 + \theta b_2$ and associated survival probability $\theta$. In this case there is only one exemption from liquidation following consecutive defaults.

4.4.2 High aggregate loss intensity

Suppose in contrast that $1 + \dot{v}_3(b_3 + b_2) > r/\lambda_4$. Then there exists $\overline{\theta}_4$ between 0 and 1 such that

$$1 + \dot{v}_3(b_3 + b_2 + \overline{\theta}_4 b_1) = \frac{r}{\lambda_4}.$$ 

In this case probation consists of two intervals. In the first, $[b_4 + b_3, b_4 + b_3 + b_2]$, there is only one exemption from liquidation following consecutive defaults. Should a loss occur servicers enter stochastic liquidation of $i = 3$ with survival probability $\theta = (u - b_4 - b_3)/b_2$. In the second, $[b_4 + b_3 + b_2, \gamma_4]$ with target defined as $\gamma_4 = b_4 + b_3 + b_2 + \overline{\theta}_4 b_1$, there are two exemptions from liquidation following consecutive defaults. Investors want the performers in that interval to fall in the probation interval of state 3, where one more default is allowed without risk of immediate liquidation.
By the concavity of investors’ continuation utility $\overline{\theta}_4 < \overline{\theta}_3$ whenever $r > 0$. In the advent of default in the bliss state, servicers find themselves within probation and do not get payments for some time. The targets consistent with servicers being paid when $i = 4$ and $i = 3$ are

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

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

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

Servicers lose from the bliss state of $i = 4$ to that of $i = 3$, not only in absolute terms, but also in relative terms.

The intuition behind these two cases is the following. Aggregate loss intensity is maximal for $i = 4$ when “contagion” has spread with the highest possible individual risk ($\alpha_4 = \alpha_3$). Looking forward, servicers’ 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 when $i = 4$, investors design probation with two intervals, meant to be reduced to a single one when $i = 3$. As shown above, the size-adjusted target, and consequently the payment made, are higher. So the size-adjusted reservation rent is improved but the size-adjusted reward received under bliss also curtailed in the advent of default.

If on the contrary delinquencies are not likely when there are four loans ($\alpha_4 \ll \alpha_3$), servicers enjoy a high rent in that state and are undermined by the downsizing. There is no need to shrink probation going forward. With only one exemption from liquidation, servicers have to wangle their way into both the stochastic liquidation and probation intervals of $i = 3$ in the advent of default. In this still rather special case it is not possible to shrink probation below a single interval since probation should not be empty. This need not be the case for higher order states with a large number of probation intervals. Thus looking forward, contraction is by one interval at most, but expansion can be sizable. The punchline is that, looking
forward, the cost of performance \( r/\lambda \) rises slowly during spells of constant individual risk and small cuts in the number of exemptions from liquidation in the bliss state hold servicers’ rent in check. In contrast, the cost of performance dwindles following a sharp worsening of credit risk and lump increases in the number of exemptions help restore their incentives to service diligently.

### 4.5 General case

We can now state the following.

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

\[
\max_{\delta_t(\cdot) \geq 0} \left\{ (ru_i + \lambda_i b_i - \delta_t) \dot{v}_i(u) + i\mu - \lambda_i (v_i(u) - \theta v_{i-1}(u - b_i)) \right\} = 0
\]

subject to

\[
du + \delta_t \, dt = ru \, dt - (b_i \wedge (u - b_{i-1})) \left( \sum dN_i^t - \lambda_i \, dt \right) - b_{i-1} (dM_t - (1 - \theta) \lambda_i \, dt) \]

\[
\theta = \frac{u - b_i}{b_{i-1}} \wedge 1
\]

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

\[
\gamma_i = \sum_{j=0}^{l(i)} b_{i-j} + \beta_i b_{i-(i-1)-1}, \quad \beta_i \in [0, 1],
\]

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 \( \beta_i \) is the probability of survival after \( l(i) + 1 \) joint defaults in the bliss state, with \( \beta_{i+1} \leq \beta_i \) if \( l(i + 1) = l(i) + 1 \) (strict inequality if \( r > 0 \)) and \( \beta_1 = 0, \beta_2 = 1 \). The cut-off rule and scale \((l(i), \beta_i)\) are uniquely determined by the recursive conditions

\[
1 + \dot{v}_i \left( \sum_{j=1}^{l(i)} b_{i-j} + \beta_i b_{i-(i-1)-1} \right) = \frac{r}{\lambda_i}.
\]

In particular \( l(i) = i - 2 \) and \( \beta_i = 1 \) 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 outstanding \( i = I - N \), and past performance, reflected in servicers’ informational rent \( u \). The minimum rent consistent with diligent servicing is \( b_i \). Given track record \( u \geq b_i \), it makes sense for investors to encourage servicers to improve their credentials before making payments. To keep them participating, they let the rent grow at a rate consistent with pool size and rate of impatience. Proposition 2 determines how far the target \( \gamma_i \) is away from \( b_i \). Once the target is reached, servicers are 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 pool 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 pool from liquidation for up to \( l(i) \) consecutive defaults, servicers’ 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 servicers can withstand without fearing liquidation under the best track record.

Under the optimal plan the level of commitment is contingent on servicers’ past performance. If \( u \) is 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 \( k \)th bucket, \( k - 1 \) exemptions are granted and the \( k \)th is reneged with some probability. The process is interrupted when \( u \) hits \( \gamma_i \) and servicers are paid. In the worst case scenario (which happens with probability zero), \( l(i) \) defaults knock the pool in one stroke and servicers’ 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 pool 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 arises when \( r = 0 \) and \( l(i) = i - 2 \). Since it makes no sense to defer payments when servicers are infinitely farsighted, investors lose an instrument and are better off letting servicers hang on to their target \( \sum_{j \leq i} b_j \) anyway. There is no risk of private benefit diversion since they enjoy 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 servicing costs ex ante. Introducing deterministic liquidation when \( r \geq 0 \) would not qualitatively change our results, as the recursive solutions would simply start from a prespecified level.

Impatient servicers are given a less ambitious target. If underlying individual risk is expected to remain constant for some time, aggregate risk is declining and the number of exemptions shrinks as losses unfold. This does not take servicers far away from target, since target and utility are reduced jointly following a loss, but lowers payments in size-adjusted terms. If underlying risk is expected to deteriorate, exemptions can be reset to a higher number. This worsens servicers’ position relative to target, but improves their prospects if lucky enough to earn their way out of trouble.

5 Implementation

The optimal contract can be implemented with tranching under realistic assumptions. We consider only the relationship between the sponsor/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 securitiesto outside investors. The sponsor seeks to maximize profits and, with given portfolio size, minimizes the amount of capital needed. Its program at time 0,

\[
\max_{u \geq b_t} \quad u - K \\
\text{s.t.} \quad K \geq I - v_I(u),
\]

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

\[\text{In the model, affiliated servicers and sponsor are treated on a consolidated basis. There is no agency problem between them.}\]
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 there are enough funds to start with, namely \( K = I - v^* \). The sponsor initiates an ABS transaction by selling the pool 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.

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 penalty \( b_i = B/(\alpha_i) \) 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. Let \( b_{L,U} \) be the common penalty in that tranche. The protection embedded in a portfolio of tranches with "optimal" weights adjusted to the underlying penalties,

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

is just the sum of penalties \( b_j \) when \( j \) runs the gamut from full size \( j = I \) to last size \( j = I - N + 1 \). Thus, the default-contingent exposure \( P(N) \) cumulates penalties inflicted on servicers since the beginning and rises from zero to \( \sum_{j=1}^{I} b_j \) when the last loan is terminated. Its variation from size \( i \) to \( i - 1 \) is driven by servicers’ constant risk shares\(^9\) \( b_i/\mu \) in the tranches. Over the first-loss piece \( b_I/\mu \) is close to one and servicers take the brunt of the losses to protect investors. Over the senior tranche, they are less exposed to default risk and a larger fraction of losses is passed through to investors. According to the optimal contract, servicers keep sharing in the risk at a declining rate, until liquidation. This is in marked contrast with standard practice

\(^9\)Because only monitored finance is viable, the gains from diligent servicing \( \mu (e\lambda_i) \) are always larger than the private benefits from shirking \( Bi \). Thus \( b_i = B/(\alpha_i) < \mu \).
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 in tranches, optimal risk prevention policy can be implemented with tranching:

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

(ii) The sponsor buys CDO tranches, weighted \( b = B/(\alpha) \), 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 \( Bi/\epsilon \) and accrued interests (rate \( r \)) are credited to the reserve account;

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

- Excess cash triggers payment to servicers;
- 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 servicers by subordinating cashflows to their 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, it writes protection by buying CDO tranches (CDS style\(^{10}\)) to match servicers’ optimal 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 sponsor were considered as a simple arbitrageur operating in the credit derivatives market. Third, the servicing and rent-preserving fees always remain at the top of the flow of funds, whether they are directly remitted to servicers or serve to replenish the reserve account.

Finally, the trust monitors performance continuously by peeking at the cash position within prescribed limits. Should it 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

---

\(^{10}\) 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).
liquidation is warranted, perhaps on the basis of her superior information. The trust always stands in for low-performing servicers, 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 sponsor maximizes profits subject to servicers conducting due diligence on delinquent borrowers and the trust breaks even.

6 Policy implications

The cost of mortgage debt has increased dramatically in recent months. Outside 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.\(^\text{11}\) 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), is the “magic” trick of incentive alignment familiar from insurance contracts. In this world, banks that originate bad loans bear the impact of losses up to a first-loss piece (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

\(^{11}\)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).
becomes a real issue. One first implication of this paper is that when continuous servicing 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. 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 should 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 when performance is poor. 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

\[12\] 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.
bank may retain the “servicing rights to exposures” without it constituting “indirect control of the exposures” and so remains 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, an originator engages in CRT without being committed to the optimal plan. It can hold fewer junior tranches and more senior tranches than necessary. Early on servicers are not penalized as they should following losses and the incentive compatibility constraint is violated. In bad states servicers receive low fees relative to the protection sold and the promise-keeping constraint breaks down. 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 secrets 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 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

---

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).
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, as the CDO premiums are waived by the sponsor except for the fraction that can be returned as rent-preserving fees to servicers.

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 inputs 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.

7 Conclusion

While the literature generally considers endogenous liquidation values with exogenously given contracts (Schleifer and Vishny, 1992), we endogenize contracts with exogenously given liquidation values. Here the emphasis is on the quality of the servicing of outstanding loans despite the fact that most attention has been paid to lending standards during origination. Diligent servicing reduces defaults on bank loans just as continuous testing of students reduces the probability of failure.\textsuperscript{14} As stressed in the Introduction, the definition of “servicers” 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.”

\textsuperscript{14}I am indebted to Robert Krainer (U. of Wisconsin) for the analogy.
Bank pay, capital requirements, securitization and liquidity risks are generally treated as areas to be regulated separately, each under its own “global framework.” The paper shows under crude assumptions that the four can be related. Servicers’ compensation must be made contingent upon past performance. Past performance is revealed to outside investors through transparent liquidity management. Liquidity management in turn is best carried out using securitization as regulatory instruments. Finally, the cost of securitization determines, up to gains on sale, the optimal capital requirement.

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 the books for as long as possible, as this would increase the income they receive from the pool, and should be constrained in the amount of time they are allowed to operate. Likewise, servicers 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 the idea that the bankruptcy code could be amended to allow for a system of graduated interventions. 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.

Liquidity regulation has a role to play to remove uncertainty about the bank’s performance and shift aggregate risk away from the institution without ruining its incentives to monitor and, as such, could be appended to solvency regulation (Tirole, 2009). Credit derivative instruments, sometimes vilified as weapons of mass destruction, can provide a first line of defense in the guaranteeing chain against systemic risk. Liquidation regulation cannot, however, provide authorities with transparency about the system’s risk exposure. As shown by Hellwig (2009), the best incentives for risk control at the level of the institution will not be able to forestall a crisis if the participants do not have information about systemic risk exposures. So far the measurement of performance depends on the trust’s ability to fulfil its obligations when needed, which is by no means a foregone conclusion if the trust is itself exposed to systemic risk. Liquidity is not meant to address systemic risk. As advocated by Tirole (forthcoming), public insurance may be needed if the guarantor is judgement-proof and the extension of liability gives rise to further externalities.
8 References


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


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


Sannikov, Yuliy, 2005, “Agency problems, screening and increasing credit lines,” Mimeo, University of California, Berkeley.


9 Appendix

9.1 Pool size analytics

9.1.1 Two loans

With normalized\(^{15}\) \(w_2^j(\theta) = b_1^{-1} v_2(u)\) the ODE specifying investors’ continuation utility when \(i = 2\) becomes

\[
\left( \frac{b_2}{b_1}(\lambda_2 + r) + r\theta \right) w_2^1 + \lambda_2 \frac{b_2}{b_1} \frac{\mu e}{B} - \lambda_2 (w_2^1 - \theta \overline{w}_1) = 0,
\]

the solution of which can be written as

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

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>(A_i^j)</th>
<th>(A_i^j)</th>
<th>(B_i^j)</th>
<th>(y_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>(\frac{b_{i+1} - b_i}{b_i} \left( A_i^{j-1} + r \right))</td>
<td>(\frac{\lambda_i}{\lambda_i - r} A_i^{j-1})</td>
<td>(\sum_{k=0}^{j-1} \frac{b_{i+k} - b_i}{b_i} \frac{\mu e}{B} + A_i^{j-1} \frac{A_i^{j-1}}{A_i^{j-1}})</td>
<td>(\frac{\mu e}{B} \left( \frac{\lambda_i - r}{\lambda_i} \frac{1}{1 - y_i} \right))</td>
</tr>
<tr>
<td>Remarks</td>
<td>(A_i^0 = \lambda_i)</td>
<td>(A_i^0 = \frac{b_i - 1}{b_i} \overline{w}_i)</td>
<td>(B_i^1 = \frac{b_i - 1}{b_i} \frac{\mu e}{B} + A_i^1 \frac{A_i^1}{A_i^1})</td>
<td>(\frac{\mu e}{B} \left( \frac{\lambda_i - r}{\lambda_i} \frac{1}{1 - y_i} \right))</td>
</tr>
</tbody>
</table>

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 \(C_2^1 = (1 + A_2^1) y_2\). When the target \(u = b_2 + b_1\) is reached, servicers receive a continuous fee of \(\delta_2 = r(b_2 + b_1) + 2B/\epsilon\) until either of the two loans defaults. Their continuation utility then jumps to \(b_1\).

A technical condition is needed to ensure that \(w_2^1\) is originally increasing in the agent’s continuation utility.

**Assumption 3** There exists \(k\) such that \(k \leq \alpha_i/\alpha_{i-1} < 1\) and \(\overline{w}_1 \equiv v_1/b_1 = \mu e/B - (\lambda_1 + r)/\lambda_1 > 1/\ln(1 + k/2)\).

\(^{15}\)The notation \(w_i^j(\theta)\) refers to a normalized solution with \(i\) loans over the \(j\)th interval \([b_i + \cdots + b_{i-(j-1)}, b_i + \cdots + b_{i-j}]\). It is defined as \(w_i^j(\theta) = v_i(u)/b_i - j\) where \(\theta\) is the position of \(u\) in that interval.
This requires that the loan revenue be all the higher in proportion of rents, the higher the maximum increase in individual risk \( \alpha_i - 1/\alpha_i = 1/k \). The condition implies that \( A_2 = \lambda_2 / (\lambda_2 - r) \bar{w}_1 > y_2 / (1 - y_2) \). Indeed, \( y_2 = (\Lambda_2^1 / (\Lambda_2^1 + r))^{\lambda_2 / r - 1} \) and \( \Lambda_2^1 = (b_2 / b_1) (\lambda_2 + r) \) imply

\[
y_2 = \left( \frac{x + 1}{x + 1 + b_1/b_2} \right)^{x-1} \leq \left( \frac{x + 1}{x + 1 + k} \right)^{x-1}
\]

where \( x = \lambda_2 / r \geq 1 \), so that

\[
\frac{y_2}{\lambda_2} \frac{\lambda_2 - r}{1 - y_2} \leq \frac{x - 1}{x} \frac{x + 1}{x + 1 + k}^{x-1} \leq -\ln \frac{2}{x + k} \leq \bar{w}_1,
\]

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

\[
\bar{w}_2 \equiv w_2^1(0) = b_2 \mu c_b / b_1 B + A_2^1 \frac{\Lambda_2^1}{\lambda_2} (A_2^1 - C_2^1) > b_2 \mu c_b / b_1 B,
\]

a result needed just below.

### 9.1.2 Three loans

In the stochastic liquidation interval \([b_3, b_3 + b_2]\) investors' normalized 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 \( C_3^1 < (1 + A_3^1) y_3 \). But we have just seen above that \( (\lambda_3 - r) / \lambda_3 A_3^1 = b_1 / b_2 \bar{w}_2 > \mu c_b / B \) and, using the same logic as in the two-loan case under Assumption 3, \( \mu c_b / B > (\lambda_3 - r) / \lambda_3 y_3 / (1 - y_3) \). It follows that \( (1 + A_3^1) y_3 < A_3^1 \) and \( \dot{w}_3^1(0) = A_3^1 - C_3^1 > 0 \).
Moreover
\[ \bar{w}_3 \equiv w_3(0) = \frac{b_3 \mu c}{b_2 B} + \frac{A_3^1}{\lambda_3} (A_3^1 - C_3^1) > \frac{b_3 \mu c}{b_2 B}, \]
a property preserved by induction across all stochastic liquidation intervals of higher order states.

The bank’s position in the second interval \([b_3 + b_2, b_3 + b_2 + b_1]\), \(\theta = (u_3 - b_3 - b_2) / b_1\), is the probability of survival given default that the pool faces after a loss. The solution is
\[
\begin{align*}
    w_3^2(\theta) &= B_3^2 + A_3^2 \theta + \phi_3^2(\theta) - C_3^2 \frac{A_3^2}{\lambda_3} \left( \frac{A_3^3 + r \theta}{A_3^3} \right) ^ {\lambda_3 / r}, \\
    \left( A_i^j + r \theta \right) \dot{\phi}_i^j(\theta) &= \lambda_i \left( \phi_i^j(\theta) - w_{i-1}^{j-1}(\theta) + B_{i-1}^{j-1} + A_{i-1}^{j-1} \theta \right), \quad \phi_i^j(0) = 0.
\end{align*}
\]
The pasting condition
\[
\dot{w}_1^3(1) = \dot{w}_2^3(0) \iff A_3^1 - C_3^1 / y_3 = A_3^2 - C_3^2 + \dot{\phi}_3^2(0)
\]
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 \( \bar{\theta}_3 \) be the upper boundary of the probation interval. If \( \bar{\theta}_3 \in (0, 1) \), it solves the system
\[
\begin{align*}
    \dot{w}_3^2(\theta) &= -1, \\
    \ddot{w}_3^2(\theta) &= 0.
\end{align*}
\]
Differentiating the ODE defining \( \phi_3^2 \) to eliminate \( \ddot{\phi}_3^2 \), one finds after some substitutions that the critical level \( \bar{\theta}_3 \) satisfies
\[
1 + \dot{w}_3^2(\theta) = \frac{r}{\lambda_3}
\]
By construction the slope of the objective function declines to \(-1\) until \( \theta = 1 \) so \( \bar{\theta}_3 \) is certainly less than one. Since \( \dot{w}_2^3(0) \) is positive \( \bar{\theta}_3 \) is away from zero. All parameters are then recovered from the boundary condition \( 1 + \dot{w}_3^2(\bar{\theta}_3) = 0 \).
9.1.3 Four loans

With \( i = 4 \) the normalized solution in the second interval \([b_4 + b_3, b_4 + b_3 + b_2]\) is

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

If probation is contained in this second interval, let \( \overline{\theta}_4 \in (0, 1) \) be the upper boundary. It is again determined by the pasting conditions \( \dot{w}_4^2 = -1 \) and \( \ddot{w}_4^2 = 0 \) leading to

\[
 1 + \dot{w}_4^1(\theta) = \frac{r}{\lambda_4}.
\]

The social value of performance is \( 1 + \dot{v}_3(u - b_4) \) one step ahead. The current relative cost of performance is \( r/\lambda_4 \). Whether or not \( \overline{\theta}_4 < 1 \) depends on which interval one step ahead has a social value of performance equal to the current cost. If this happens when \( v_3(u) = b_2 w_4^3(\theta) \), the stochastic liquidation interval of regime 3, condition (4) is met in the second interval of regime 4 and yields the target \( \gamma_4 = b_4 + b_3 + \overline{\theta}_4 b_2 \). If this happens when \( v_3(u) = b_1 w_3^2(\theta) \), the probation interval of regime 3, condition (4) cannot be met because the social value of performance is still high relative to its cost at \( b_3 + b_2 \). But then \( 1 + \dot{v}_3(b_3 + b_2) > r/\lambda_4 \) and the smooth pasting condition for the boundary \( \overline{\theta}_4 \) in the third interval \([b_4 + b_3, b_4 + b_3 + b_2 + b_1]\) reads

\[
 1 + \dot{w}_5^2(\theta) = \frac{r}{\lambda_4}
\]

and yields the target \( \gamma_4 = b_4 + b_3 + b_2 + \overline{\theta}_4 b_1 \).

In either case, the remaining unknown parameters can be recovered by continuity from the slope of investors’ utility \( v_4 \), starting from the outermost interval with \( \dot{v}_4(\gamma_4) = -1 \) and working back to the left.
10 Proof of Proposition 3

From servicers’ integrated promise-keeping constraint (2) along the optimal path, we know that for all \( t \leq \tau \)

\[
u_t = u^* + \int_0^t \left( ru_s + \frac{Bi}{e} \right) 1_{\{u_s < \gamma_i\}} \, ds - \int_0^t b_i \wedge (u - b_{i-1}) \sum_j dN^j_s - \int_0^t b_{i-1} \, dM_s,
\]

where \( i = I - \sum_j N^j_t \). By construction, the protection sold by the sponsor is

\[
P(N_t) = \int_0^t b_j \sum_j dN^j_s + \left( \sum_{j<i} b_j \right) M_{t\wedge \tau}
\]

since at \( t = \tau \) the default count jumps from \( \sum_j N^j_t \) to \( I \). Thus

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

where the martingale

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

is the trust’s cumulated cost resulting from intervention after stochastic liquidation. With probability \( \theta \), the pool is maintained and the trust pays the shortfall \( \Delta \xi = b_i + b_{i-1} - u \). With probability \( 1 - \theta \) the pool is liquidated and the trust wins the residual balance \( -\Delta \xi = u - b_i \). Evaluating (5) at \( t = \tau \) with \( u_\tau = 0 \) and \( N_\tau = I \), we get

\[
\sum_{j \leq i^*} b_j = u^* + \int_0^\tau \left( ru_s + \frac{Bi}{e} \right) 1_{\{u_s < \gamma_i\}} \, ds + \xi_\tau + \sum_{j<i^*} b_j \quad (6)
\]

where \( i^* = I - \sum_j N^j_t \) is the pool size at liquidation.

The sponsor maximizes its profit since by construction

\[
u^* = E \int_0^\tau e^{-rt} \left( r\gamma_i + \frac{Bi}{e} \right) 1_{\{u_t = \gamma_i\}} \, dt
\]

\[
= E \int_0^\tau e^{-rt} \delta_t \, dt.
\]
The trust’s costs and benefits in the course of the relationship are as follows

<table>
<thead>
<tr>
<th></th>
<th>Cost</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t = 0$</td>
<td>$u^* + v^*$</td>
<td>—</td>
</tr>
<tr>
<td>$(t, t + dt)$</td>
<td>$(ru_t + \frac{B_i}{\tau}) \mathbf{1}_{{u_t &lt; \gamma_i}}$</td>
<td>$i\mu - \delta t + \Sigma_t$</td>
</tr>
<tr>
<td>Liquidation</td>
<td>$\xi_f + \sum_{j &lt; i} b_j$</td>
<td>—</td>
</tr>
</tbody>
</table>

where $\Sigma_t$ 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

$$v^* = E \int_0^\tau (i\mu - \delta_t) \, dt$$

$$\sum_{j \leq I} b_j = E \int_0^\tau \Sigma_t \, dt,$$

the first equality by design, the second by arbitrage since at date $\tau$, the pool is liquidated and $N_\tau = I$. The trust breaks even.