DOCUMENT

DE TRAVAIL

N° 377

BANK MONITORING INCENTIVES AND OPTIMAL ABS

Henri Pagès

April 2012

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

BANK MONITORING INCENTIVES AND OPTIMAL ABS

Henri Pagès

April 2012

Les Documents de travail reflètent les idées personnelles de leurs auteurs et n'expriment pas nécessairement la position de la Banque de France. Ce document est disponible sur le site internet de la Banque de France « www.banque-france.fr ».

Working Papers reflect the opinions of the authors and do not necessarily express the views of the Banque de France. This document is available on the Banque de France Website “www.banque-france.fr”.

Bank monitoring incentives and optimal ABS

Henri Pagès*

*Forthcoming in the *Journal of Financial Intermediation*, http://dx.doi.org/10.1016/j.jfi.2012.06.001. Tel: +33 142922999, Fax: +33142925989, Email: henri.pages@banque-france.fr. I am indebted to Bruno Biais for pointing out a flaw in an early version and to Denis Gromb and Jean-Charles Rochet for their invaluable guidance and support. I would like to thank two anonymous referees, Hendrik Hakenes, Christian Hellwig and George Pennacchi for insightful comments and suggestions, Monique Jeanblanc and Dylan Possamaï for their help with the mathematics and simulations, as well as discussants and participants in many seminars and conferences over the last three years. All errors are mine. The usual disclaimer applies.

Abstract

The paper examines a continuous-time delegated monitoring problem between competitive investors and an impatient bank monitoring a pool of long-term loans subject to Markovian "contagion." Moral hazard induces a foreclosure bias unless the bank is compensated with the right incentive-compatible contract. Fees are paid when the bank's performance is on target and liquidation arises when the bank's performance is sufficiently poor. I show that the optimal contract can be implemented with a whole loan sale involving both credit risk retention based on ABS credit default swaps and credit enhancement in the form of a reserve account. The optimal securitization bears out rulemaking recently proposed in the wake of the Dodd-Frank Act on a number of controversial provisions. I argue that further efficiency gains could be reaped by extending the role of the "premium capture" account into a liquidity buffer capturing performance-based compensation as a way of increasing skin in the game over the life of the deal.
Keywords: ABS Credit Default Swaps, Banking Regulation, Default Correlation, Dynamic Moral Hazard, Optimal Securitization, Risk Retention.

JEL Classification: G21, G28, G32.

Résumé

Le papier examine un problème de surveillance déléguée en temps continu entre des investisseurs compétitifs et une banque impatiente, gestionnaire d'un lot de prêts à long terme sujets à « contagion » Markovienne. L'aléa moral induit un biais de saisie, à moins que la banque ne soit dédommagée par un contrat suffisamment incitatif. La commission est versée lorsque la performance est conforme aux attentes et la liquidation ordonnée lorsqu'elle est insuffisante. Je montre que le contrat optimal peut être mis en vigueur par une vente complète associant rétention du risque avec CDS d'ABS et rehaussement du crédit à travers un compte de réserve. La titrisation optimale est en accord avec la réglementation récemment inspirée du Dodd-Frank Act sur certains points litigieux. J'avance que des gains supplémentaires d'efficience pourraient être retirés en conférant au compte de « capture de prime » un rôle élargi de réserve de liquidité, permettant d'assujettir la rémunération à la performance et d'accroître la part du risque retenu jusqu'au terme de la transaction.
Mots-clés : ABCDS, aléa moral dynamique, corrélation des défauts, réglementation bancaire, rétention du risque, titrisation optimale.

Codes JEL: G21, G28, G32.

1 Introduction

The home loan crash has exposed the conceptual weaknesses of securitization agreements originated in the first years of the last decade and prompted the US government to impose tight deadlines for the adoption of regulations for credit risk retention. The Dodd-Frank Act[1] prescribes that sponsors retain at least five percent of the credit risk in most securitization transactions. In April 2011, six federal agencies (Agencies) issued a Notice of Proposed Rulemaking (NPR), including a request for comments on the requirements for and exemptions from such risk retention. They have received about 13,000 comments, suggesting that the "skin in the game" requirements remain one of the most controversial issues in the Dodd-Frank Act.

A helpful approach at this juncture would be to explicitely examine the optimality of the NPR provisions in a world with optimal contracts. One disputable feature is the "one-size-fits-all" threshold of five percent, which has little economic basis in the light of the sheer variety of asset classes and structures used in securitization.[2] Beyond this standard criticism, however, there are other features that warrant more scrutiny.

First, the options sponsors have in slicing their Asset-Backed Securities (ABS) have different implications for their choice of monitoring effort in the time dimension.[3] With "horizontal" risk retention, the sponsor has residual rights through the equity tranche. Compensation is then related to performance, because high losses can interrupt the payments made to the residual holder. On the other hand, the equity tranche entitles the sponsor to capture the substantial premium that is typical of subprime securitizations and provides poor incentives to monitor if exhausted. In contrast, a "vertical" slicing in the form of a share of all ABS issued maintains skin in the game throughout, but is inconsistent with the suspension of payments that a poor performance might require. It is doubtful that the "L-shaped" risk retention, a combination of the two sub-optimal schemes above, would suggest itself as the optimal incentive alignment device.

Second, and more importantly, there is no distinction in the regulation between the credit risk and the economic interest retained in a securitization. At any point in time, the balance of ABS interests specifies

[1] Section 941 (2010). At the end of the same year, the Committee of European Banking Supervisors issued final guidelines on application of a package of amendments known as the Capital requirements Directive II, with similar provisions concerning securitization risk retention in the European Economic Area.

[2] See Board of Governors of the Federal Reserve System (2010).

[3] See relatedly Fender and Mitchell (2009) in a setting of adverse selection.

both the credit risk retained by the sponsor and the monetary benefits it receives. This constitutes a stumbling block in the implementation, as the Agencies would like sponsors to increase their risk retention while the industry objects that this would raise the cost of securitization. The "premium capture" provisions are a case in point. The rule would effectively prohibit sponsors from realizing value at closing by requiring them to deposit cash into a reserve account over and above the five per cent risk retention requirement. Given the fierce opposition[4] voiced by industry, it is likely that the Agencies will move forward with a modified proposal for rulemaking.[5]

The purpose of this paper is to study optimal securitization when the sponsor remains involved with its retail originations and can engage in unobservable actions that result in private benefits at the expense of performance. The assumption that banks impact the underlying riskiness of the pool over time is a metaphor for the distinction between its exogenous base quality and the endogenous default probability that obtains after monitoring. Given competitive investors, the goal is to solve a security design problem by finding out which compensation scheme maximizes the sponsor's payoff subject to a zero-profit condition for investors and an incentive compatibility constraint for monitoring.

What distinguishes the paper from earlier analyses is the view that moral hazard is a dynamic rather than one-shot problem. While weak underwriting standards have played a major role[6] in the subprime-mortgage financial crisis, the Dodd-Frank Act focuses on key improvements[7] that should assist market participants in coping with moral hazard at origination. The assumption of continuous moral hazard seems the correct one because sponsors are compelled to hold unhedged risk exposures over a period of time that may extend to forty years, which would be hard to justify[8] were risk retention solely designed to prevent the poor origination of credit. The assumption resonates well with the Federal Register's (2005) statement that the role of servicing in ABS transactions is as important to the performance of the pool as its initial composition and characteristics.

[4] See Financial Services Roundtable (2011), Securities Industry and Financial Markets Association (2011).

[5] See Myles (2012).

[6] See for instance Dell'Ariccia et al. (2009), Demyanyk and Van Hemert (2011), International Monetary Fund (2008), Keys et al. (2010), O'Keefe (2009).

[7] See in particular Sections 942 (disclosure requirements), 943 (representations and warranties), 945 (due diligence analysis).

[8] Hartman-Glaser et al. (2012) show that the timing of cashflows accruing to the bank is key in a one-shot screening problem, but that the maturity of the optimal contract is likely to be very short even with long-lived mortgages.

There is much that an originating bank can do to improve performance over the life of a transaction. First, quality control processes using continuous flow of information[9] help lenders exercise due diligence in evaluating borrowers' current income and keep track of those who narrowly passed the underwriting guidelines. Subprime mortgage lending is an exception-laden process. Second, servicers can efficiently assist troubled borrowers by acting early and firmly and extracting concessions before mortgages become seriously delinquent. How sponsors select, compensate and discipline servicers can affect loss severity by as much as 30% according to standard estimates.[10] Agarwal et al. (2012) find evidence of significant systematic changes in the delinquency rates of state-chartered banks' real estate loans, following a "rotation" policy between federal and state supervisors at predetermined time periods. This suggests that banks are able to undertake corrective actions in the event of delinquencies, when they have incentives to do so. For example, the servicer may advance unpaid interest, keep tax and insurance bills current and manage delinquencies in order to mitigate the risks of borrowers "walking away" from their mortgages.

The contribution of the paper is twofold. First, I characterize the optimal contract between an impatient bank and deep-pocket investors given a pool of identical "perpetuities" with constant cash flows and a liquidation value of zero. The optimal contract is a dynamic version of the delegated monitoring problem studied in the relationship services literature, where the assumption about default is intermediate between the complete diversification in Diamond (1984) and the perfect correlation in Holmström and Tirole (1997). As in the static borrower-lender relationship of Innes (1990), the best way to provide correct incentives for effort is a carrot-and-stick approach giving the bank maximum reward when performance is good and maximum penalty when performance is poor. The difference in the dynamic setting is that there is an intermediate case, when performance is below target yet does not vindicate liquidation of the pool. Payments are always suspended or delayed further following defaults. The length of the "probation" period reflects the influence that the bank keeps through its monitoring. Performance-based compensation actually ensures that the bank retains a stake in the pool over time and, as in Besanko and Kanatas (1993), is able to commit to monitoring under external capital market financing.

[9] See Dichev and Skinner (2002), Norden and Weber (2010).
[10] See Fitch (2003), Moody's (1999).

Second, I show how a specific securitization scheme decentralizes the optimal contract. The implementation takes the form of a whole loan sale with monitoring retained. The sponsor does not hold a residual interest in its balance sheet for the right to receive a portion of future cash flows, but credit enhances the deal with a cash reserve account. The reserve account serves both as a buffer for protection extended in the form of ABS credit default swaps (ABCDS) and as an instrument to tie the amount and timing of compensation to performance. Credit enhancement is known to play an important role in loan securitization.[11] In Greenbaum and Thakor (1987), the optimum level of credit enhancement allows for the truthful revelation of credit risk, which helps the bank save on screening costs and accept a higher level of risk than if it were funded with deposits. In my model, the reserve account reveals the level of underlying performance, which reduces the rent of the monitoring bank and allows it to retain risk at a lower cost than if it were funded with deposits.

The optimal securitization scheme can be contrasted with the NPR provisions. I find support for features that the securitization industry proved sharply critical of, in particular the establishment of a "premium capture" account, the absence of "sunset" provisions regarding long-term unhedged exposures to credit risk and the risks stemming from a potential reclassification of securitization transactions as a consolidation in sponsors' balance sheets. Perhaps the most comforting conclusion from the analysis is that the five percent credit risk retention required by the Dodd-Frank Act can be collateralized at a low funding cost, using performance-based compensation as a means of freeing resources that are earmarked in the reserve account to increase credit support.

The paper belongs to the recent and fertile theory on dynamic moral hazard, as illustrated by Biais et al. (2007), DeMarzo and Fishman (2007a, 2007b), DeMarzo and Sannikov (2006) or Sannikov (2008). The setting is closely related to the continuous-time principal-agent model with large and infrequent risks studied by Biais et al. (2010), with the twist that (i) ABS refer to a discrete pool of assets that self-liquidate[12] and (ii) defaults are imperfectly correlated according to a Markovian view of "contagion," in which the underlying

[11] Similarly, Gorton and Pennachi (1995) show that collateralization can be a securitized banking activity in which the bank continues to provide special credit evaluation or monitoring services of the loans backing the securitized products it provides.

[12] See Federal Register (2005).

default intensities cannot decrease with the arrival of new defaults. In Biais et al. (2010), moral hazard is about large and infrequent risks. As in my model, investors inflict reductions in the agent's continuation utility following losses and downsizing is inefficient in the first-best but necessary in the second-best when performance is poor. My analysis offers a first description of unpredictable downsizing in a non-stationary context. Technically, the extension is non-trivial for two reasons. First, I have to rule out a set of parameter values such that the incentive compatibility constraint is not binding when performance is poor. Second, I find a sufficient condition under which the best way to provide the bank with incentives to monitor when perfomance is poor is the threat of termination rather than any other form of random downsizing.

There is compelling evidence about the importance of dynamic moral hazard in securitization. Ashcraft and Shuermann (2008) provide insights into the frictions that arise in the atomized setting of securitization. Piskorski at al. (2010) show empirically that securitization induces a foreclosure bias in private subprime mortgages. This contrasts with Adelino et al. (2009), who find that servicers do not use modifications as a frequent renegotiation tool. Gan and Mayer (2006) find that when servicers hold the first-loss piece, they appear to behave more efficiently, with a positive impact on the price of junior tranches. In Cantor and Hu (2007), the weaker performance of certain types of sponsors is related to their incentives to economize on quality servicing. Pennington-Cross and Ho (2006) examine the heterogeneity of servicers in securitized subprime mortgages and estimate large variations in loan default probabilities relative to a reference group.

The paper proceeds as follows. Section 2 presents the model and assumptions. The analysis starts in Section 3 with the formulation of the incentive compatibility and limited liability constraints and goes on with their implication for the payoff possibility frontiers under the optimal contract. Section 4 deals with implementation issues. Section 5 provides a discussion of the proposed rulemaking regarding risk retention requirements by securitizers. Section 6 concludes. Sketches of proofs and robustness results are presented in the Appendix. A formal derivation of the analytical results in Section 3 is available in a companion mathematical paper (Pagès and Possamaï, 2012).

2 The model

2.1 Continuous monitoring technology

Consider a model with universal risk neutrality in which time is continuous and indexed by $t \in [0, \infty)$. The risk-free interest rate is normalized to zero. At $t = 0$, a bank owns a claim to a pool of I one-dollar loans indexed by $i = 1, \ldots, I$ which are ex ante identical. Each loan is modeled as a defaultable perpetuity yielding a continuous interest payment μ until it defaults. The model focuses on credit risk and abstracts from distributions of principal arising from amortization and prepayments. Once a loan defaults, it gives no further payments. The infinite maturity and no-recovery assumptions are made for tractability.

Let $N_t = \sum_{i=1}^{I} 1_{\{\tau^i \leq t\}}$ be the default count at time t, where τ^i denotes the default time of loan i. At any time t, the pool balance is the number of outstanding loans $I - N_t$. Since all loans are a priori identical, they can be reindexed after defaults. The bank decides whether it monitors loan $i = 1, \ldots, I - N_t$ (action $e_t^i = 1$) or not ($e_t^i = 0$) at each point in time. Monitoring is costly, as the bank enjoys a flow of private benefits B per non-defaulted loan that is not monitored. These private benefits capture the opportunity cost of various wasteful activities the bank can indulge in when shirking. It is natural to assume that these benefits depend on the number of loans outstanding, since a larger pool requires more monitoring.

The rate at which loan i defaults is controlled by its hazard rate α_t^i, specifying the loan's instantaneous default probability conditional on the history up to time t. Individual hazard rates are assumed to depend both on the bank monitoring choice and on the pool balance. Specifically, the hazard rate of (non-defaulted) loan i at time t is

$$\alpha_t^i = \alpha_{I-N_t} \left(1 + (1 - e_t^i)\epsilon\right), \tag{1}$$

where the parameters $\{\alpha_j\}_{1 \leq j \leq I}$ represent common individual "baseline" risk under monitoring when the pool balance is j and ϵ is the relative impact of shirking on default risk.[13] Note that as per Equation (1), monitoring affects risk only at the time it is exerted. Letting α_t^i depend on the pool size is a way of modelling

[13] One should write $\alpha_{I-N_{t-}}$ instead of α_{I-N_t} for the predictable intensity. I refrain from the cumbersome notation except in places where it clarifies the exposition.

imperfect correlation across default times. Correlation has been shown to be important for the distribution of losses among tranches[14] and plays an important role for the design of optimal securitization in the analysis. As in models recently introduced in the credit field,[15] default correlation is here induced by contagion effects, i.e., individual defaults can cause a jump in the default intensity of the remaining loans. The formulation is the simplest Markovian setting one can define, where the underlying individual default intensity is actually a Markov chain with state space given by the pool size.

From Equation (1), if the bank fails to monitor $k_t \equiv \sum_{i=1}^{I-N_t} \left(1 - e_t^i\right)$ loans, aggregate default intensity is

$$\lambda_t^k = \alpha_{I-N_t} \left(I - N_t + k_t \epsilon\right). \tag{2}$$

I call k the shirking intensity. When $k = 0$, the aggregate default intensity λ_t^0 can be written simply as λ_{I-N_t} where $\lambda_j \equiv j\alpha_j$. One interpretation of $k_t \epsilon \alpha_{I-N_t}$ is the *foreclosure bias*[16] due to poor ex-post monitoring when the bank lacks the incentives to produce an efficient level of monitoring.

At any time, individual loans can be liquidated, i.e., removed from the pool. The liquidation value is normalized to zero. This illiquidity is meant to reflect the bank's advantage in handling soft information and dealing with information-problematic borrowers vis-a-vis less capable outsiders. In the securitization of subprime mortgage loans, for instance, an issuer directly involved in its retail originations can maintain special relationships with its borrowers, understand their activities, set constraints and know best how to work on those behind on payments. For simplicity, any third party that acquires the loans is unable to collect cash flows.

2.2 Investors

The bank can fund the pool internally at a cost $r > 0$. Internal funding costs reflect the bank's limited access to capital or deposits and may include any regulatory or agency costs associated with this source

[14] See for example Azizpour et al. (2011), Duffie (2008), Franke and Krahnen (2006).
[15] For example Davis and Lo (2001), Frey and Backhaus (2008), Jarrow and Yu (2001) or Yu (2007).
[16] Piskorski et al. (2010) define the foreclosure bias as the difference in foreclosure rates observed when loans are held by the bank compared to similar loans that are securitized.

of financing. The bank can also raise funds from a competitive investor that values income streams at the prevailing risk-free interest rate. As the bank is more impatient than investors, gains from trade are realized whenever the former sells claims to remote cash flows to the latter. The positive discount rate plays a role similar to risk aversion in a typical moral hazard problem and simplifies the analysis. Both the bank and the investor observe the history of defaults and liquidations.

Assumption 1 *Loans have positive net present value under monitoring, i.e., if $\bar{\alpha}_I$ is the harmonic mean of $\{\alpha_i\}_{1\leq i\leq I}$,*

$$\mu > \bar{\alpha}_I. \tag{3}$$

Under monitoring, the expected duration of the next-to-default loan in a pool of j loans is $1/\lambda_j$. Hence, the average revenue from the pool is μ/α_j over that period, of which $1/I$ is ascribed to the original loan. Summing over j, one finds that the payoff of a loan is $\mu/\bar{\alpha}_I$, where $1/\bar{\alpha}_I$ is the expected lifetime of a loan at inception. This must be above the initial unit cost for the loan to be worth making under monitoring.

Assumption 2 *Monitoring is efficient,*

$$\mu - r\frac{B}{\epsilon\bar{\alpha}_I} > \frac{\mu + B}{1+\epsilon}. \tag{4}$$

The rationale behind the efficiency condition (4) is as follows. Providing the bank with incentives to monitor is costly when its discount rate is positive. The bank's non-pledgeable income flow due to monitoring is B/ϵ per loan. Given that the cost of carrying a stake in the loan for incentive purposes is r, the left-hand side is simply the net expected payoff from a loan if it is monitored from balance I onward. Alternatively, the bank can sell the loan and not monitor it. The probability of default increases by ϵ when the bank reaps private benefits, with a loan payoff as on the right-hand side. Private benefits appear on both sides of the inequality because, under monitoring, they are forgone by the bank and raise the cost of its stake.

The equivalent formulation

$$r\frac{B}{\epsilon\bar{\alpha}_I} < \frac{\mu\epsilon - B}{1+\epsilon} \tag{5}$$

imposes an upper bound on the bank's discount rate. It says that the discount rate should not be so large that the cost of the rent extracted by a monitoring bank outweighs the pecuniary gains stemming from the use of the monitoring technology. I verify later in the analysis that, under Assumption 2, $k_t = 0$ is (almost surely) the only efficient outcome.

Assumption 3 *Individual default risk is (weakly) increasing with past default*

$$\alpha_j \leq \alpha_{j-1}, \qquad \textit{for all } j \leq I. \tag{6}$$

Uncorrelated default risk corresponds to a constant $\{\alpha_j\}_{1 \leq j \leq I}$ and perfect correlation corresponds to $\alpha_j = \infty$ for all $j \leq I - 1$. The assumption addresses one channel of default clustering through a contagion effect related to the size of the pool. Note that the impact of default does not fade away with time, as would be the case if mean-reversion were allowed for through additional idiosyncratic processes or more involved self-exciting effects.[17] The assumption is consistent with the empirical evidence presented by Laurent et al. (2008), who estimate this simple homogeneous Markov contagion framework using a dependence structure generated by a one-factor Gaussian copula or a typical base correlation for the iTraxx. They come to the stronger conclusion that *aggregate* default risk tends to increase with past default.

2.3 Contracts

Both the bank and investors can fully commit[18] to a long-term contract. Contracts are designed and offered by investors to the bank on a take-it-or-leave-it basis at time 0. A contract determines how cash flows are shared between the bank and the investor and how loans are liquidated, if ever, depending on the history of defaults and liquidations. Without loss of generality, I specify that an investor receives all cash flows and makes transfers to the bank. The bank is protected by limited liability. Because the bank can only pledge the revenues from the pool, the cumulative transfers from the investor to the bank, $D = \{D_t\}_{t>0}$, are required to be (weakly) positive and increasing. It will be verified in Proposition 4 that transfers are absolutely continuous with respect to time under the optimal contract. In that case, the flow of transfers can be written as $dD_t = \delta_t \, dt$ with $\delta_t \geq 0$.

[17] See for example Aït-Sahalia et al. (2010), Giesecke et al. (2011).

[18] The last assumption is standard in the dynamic moral hazard literature and allows one to focus on moral hazard as the single source of market imperfection.

Liquidations could be partial, involving the removal of a state-dependent number of loans, or stochastic, implying that part or all of the loans are liquidated with state-dependent probability. I focus on a specific liquidation policy and show later when it is optimal. The contract states that all loans can be liquidated immediately after default, depending on the results of a lottery. This can be interpreted as the intervention of a trustee who conditions the liquidation on a measure of past performance to be determined contractually.

Let $\tau \leq \infty$ be the time of liquidation and $H_t = 1_{\{t \geq \tau\}}$ the corresponding liquidation indicator. The contract specifies the probability $\theta = \{\theta_t\}_{t>0}$ with which the pool is maintained given default, so that given $dN_t = 1$:[19]

$$dH_t = \begin{cases} 0 & \text{with probability } \theta_t, \\ 1 & \text{with probability } 1 - \theta_t. \end{cases} \tag{7}$$

The hazard rates associated with the default and liquidation processes N_t and H_t are λ_t^k and $(1 - \theta_t)\,\lambda_t^k$, respectively.

Each infinitesimal time interval $[t, t + dt)$ unfolds as follows:

1. $I - N_t$ loans are performing at time t;

2. The bank chooses to leave $k_t \leq I - N_t$ loans unmonitored and monitors the $I - N_t - k_t$ others, enjoying private benefits $k_t B\, dt$;

3. The investor receives $(I - N_t)\,\mu\, dt$ from the pool and pays $dD_t = \delta_t\, dt \geq 0$ as fees to the bank;[20]

4. With probability $\lambda_t^k\, dt$ defined by the aggregate equation (2), there is a default ($dN_t = 1$);

5. Given default, the pool is maintained ($dH_t = 0$) with probability θ_t or liquidated ($dH_t = 1$) with probability $1 - \theta_t$.

[19] It can be shown that, starting from an $\mathcal{F}$-predictable process θ with values in $[0, 1]$, where $\mathcal{F}$ is the filtration generated by N, one can construct an enlarged filtration $\mathcal{G}$ with stochastic time of liquidation τ verifying (7). The proof is available upon request.

[20] In principle the difference $(I - N_t)\,\mu - \delta_t$ could be negative. It will be shown that this is never the case under the optimal contract.

2.4 Informational frictions

The pool balance is affected by the bank's monitoring decision. To make this dependence explicit, let E^k denote the expectation under the probability P^k generated by the shirking intensity k over the paths of N. Given a contract (D, θ) and a shirking intensity k, the bank's and investor's current continuation utilities, or expected discounted payoffs, are respectively

$$u_t^k = 1_{\{t \leq \tau\}} E^k \left[\int_t^\tau e^{-r(s-t)} \left(dD_s + Bk_s \, ds \right) \Big| \, \mathcal{G}_t \right], \tag{8}$$

$$v_t^k = 1_{\{t \leq \tau\}} E^k \left[\int_t^\tau \left(\left(I - N_s \right) \mu \, ds - dD_s \right) \Big| \, \mathcal{G}_t \right], \tag{9}$$

where $\mathcal{G}_t$ is the information generated by default and liquidation up to t. Because shirking and liquidation are inefficient, the social optimum requires $k = H = 0$. Social surplus

$$\begin{aligned} S &= E\left[\int_0^\infty \left(I - N_t \right) \mu \, dt \right] - I \\ &= E\left[\sum_{j=1}^{I} \int_0^\infty j\mu 1_{\{I - N_t = j\}} \, dt \right] - I \quad = \quad I \left(\mu / \bar{\alpha}_I - 1 \right), \end{aligned} \tag{10}$$

is positive under Assumption 1. In the first-best, transfers boil down to a single lump-sum payment at date 0 as the bank is impatient. Moreover, as investors are competitive, the bank captures the full surplus from the project ($u_0 = D_0 = S$), and the investor breaks even ($v_0 = I$).

From now on, I assume that the bank's monitoring decision is not observable by investors and conjecture that the bank receives a flow of transfers $dD_t = \delta_t \, dt$. This leads to a dynamic moral hazard problem, as gains from trade are impaired by higher default rates when the bank shirks. Given Assumption 2, the contract (δ, θ) must use observations on defaults to give the bank incentives to monitor at all points in time.

3 Optimal contracting

A shirking intensity k is incentive-compatible with respect to (δ, θ) if it maximizes (8) at all points in time. Investors' problem is to design a contract (δ, θ) with incentive-compatible k that maximizes their current expected payoff for all t, subject to a given reservation utility for the bank:

$$\begin{aligned} v_j(u) &= \max_{\{\delta,\theta,k\}} E^k \left[\int_t^\tau \left((I - N_s)\,\mu - \delta_s \right) ds \,\middle|\, \mathcal{G}_t \right] \qquad (11) \\ \text{subject to} \quad & E^k \left[\int_t^\tau e^{-r(s-t)} (\delta_s + Bk_s)\, ds \,\middle|\, \mathcal{G}_t \right] \geq u \\ & k \text{ incentive-compatible with respect to } (\delta, \theta), \end{aligned}$$

where $j = I - N_t$ is the balance of the pool at time t. The value function $v_j(u)$ is the highest expected payoff an investor can afford given current j and u. The policy relies on two instruments: compensation δ and stochastic liquidation τ. In line with the literature on dynamic moral hazard, these decisions are made on the basis of two state variables, the balance of the pool j and the bank's expected continuation payoff u, which are sufficient statistics in this time-homogeneous set-up. While the former reflects the total number of losses, the latter summarizes the track record of performance.

3.1 Incentive compatibility and limited liability

The bank's expected payoff may change for incentive purposes following the arrival of new information. If a default occurs and the pool is maintained, payments can be suspended for some time, causing utility to fall. Should the pool be liquidated, payments can be delayed further or even called off altogether, resulting in a further utility drop. On the other hand, in between defaults, the bank should not be liable for the risks it does not control. The following promise-keeping equation states that the bank's reward must take into account the expected cost of the jumps in utility that it incurs in the event of default.

Lemma 1 *For any k, there exist predictable processes h^1 and h^2 such that the bank's continuation utility* *satisfies the promise-keeping equation until liquidation:*

$$du_t^k + (\delta_t + Bk_t)\,dt - \lambda_t^k \left(h_t^1 + (1-\theta_t)h_t^2\right)\,dt = ru_t^k\,dt - h_t^1 dN_t - h_t^2 dH_t. \tag{12}$$

Consider in turn the discontinuous and continuous parts of du_t^k. Following default ($dN_t = 1$), the bank's continuation utility jumps from u_{t-}^k to $u_t^k = u_{t-}^k - h_t^1$ if the pool is maintained ($dH_t = 0$) and from u_{t-}^k to $u_t^k = u_{t-}^k - h_t^1 - h_t^2$ if it is liquidated ($dH_t = 1$). I refer to h^1 and h^2 as the contractual *penalties* incurred by the bank given default. Otherwise there is no loss ($dN_t = dH_t = 0$) and the bank's dividend has two components. One is $\delta_t + Bk_t$, the contractual fee complemented by any private benefits that accrue as a result of shirking. The other is negative and corresponds to the expected cost of the penalties, $\lambda_t^k \left(h_t^1 + (1-\theta_t)h_t^2\right)$. The promise-keeping equation (12) states that, in the absence of jumps, the expected rate of change in the bank's continuation payoff plus the rate of dividends is equal to its discount rate r.

Lemma 2 *Given a contract (δ, θ), the zero shirking intensity is incentive-compatible if and only if for almost every $t \in [0, \tau]$*

$$h_t^1 + (1-\theta_t)h_t^2 \geq b_{I-N_t}, \qquad \text{where } b_j \equiv \frac{B}{\epsilon\alpha_j}. \tag{13}$$

For $k = 0$ to be incentive-compatible, the bank's continuation payoff must drop by at least b_j on average immediately after default when the current balance is j. If the bank chooses to monitor a loan, the expected penalty relief is $\epsilon\alpha_{I-N_t}\left(h_t^1 + (1-\theta_t)h_t^2\right)$, which must be larger than the opportunity cost B. The condition is the same irrespective of the number of loans that are monitored, as both the private benefits and the aggregate intensity of default vary in the same proportion. Hence, (13) implies that there can be no interior solutions.

Observing that the bank's continuation utility must remain positive for a going pool under the limited liability constraint, we know that $u_t \geq b_j$ on $\{I - N_t = j\}$. Otherwise, penalties could not meet the incentive compatibility condition (13). One can interpret b_j as the minimum rent consistent with monitoring j loans, the usual non-pledgeable income reflecting the attractiveness of private benefits under shirking. Two additional constraints restrict the range of admissible values for h^1 and h^2. First, the pool can be maintained immediately after default only if the bank's continuation utility does not violate limited liability

$$u_{t-} - h_t^1 \geq b_{j-1}, \qquad \text{on } \{I - N_t = j\}. \tag{14}$$

Second, the bank must forfeit any rights once the pool is liquidated. Since the pool makes no further payments, any cash flows are transfers from the investor to the bank. Given the bank's impatience, they can only take the form of a lump-sum payment immediately after liquidation. Viewed at time 0, any payment at liquidation would lower social surplus as it penalizes the investor more than it benefits the bank. Since this has no incidence on monitoring, it would be inefficient. The constraint $u_\tau = 0$ implies in turn that at all times:

$$u_{t-} = h_t^1 + h_t^2. \tag{15}$$

3.2 Pay-off possibility frontiers

As the bank enjoys limited liability over the life of the deal, its continuation utility must exceed the minimum rent consistent with monitoring b_j. Should it come to operate the last loan in the pool, however, its continuation utility would not stray above b_1. The reason is that investors are no longer concerned about the bank's performance following default, after all is lost, so that its continuation utility is maintained at the reservation level. Since $j = 1$ is a degenerate special case, it is convenient to consider the monitoring of a single loan before turning to the general case.

3.2.1 Single loan: Constant utility

It is optimal to set the bank's continuation utility at $u_t = b_1$. The bank looses its informational rent b_1 in the case of default and this is enough to keep it monitoring. The incentive compatibility condition (13) is binding and the promise-keeping equation (12) specifies a continuous payment of $\delta_t = b_1 (r + \lambda_1)$ until default τ^1. The expected payoff for the investor is:

$$\overline{v}_1 \equiv v_1(b_1) = E\left[\int_t^{\tau^1} (\mu - \delta_s)\, ds \;\Big|\; \mathcal{G}_t\right] = \frac{\mu - b_1 (r + \lambda_1)}{\lambda_1}. \tag{16}$$

It is easy to see that $\overline{v}_1 \geq 0$ under Assumption 2. What if the bank starts with a higher utility level than b_1? Because the pool value is fixed in the range $[b_1, \infty)$, the efficient payoff possibility frontier between the bank and the investor has slope of -1 and

$$v_1(u) = \overline{v}_1 - (u - b_1), \qquad u \geq b_1. \tag{17}$$

If one were to start in this region, the optimal contract would entail an immediate lump-sum payment of $u - b_1$ to the bank, counterbalanced by a drop in its continuation utility to the reservation level b_1. However, it turns out that v_1 is never used in the range $u > b_1$ under the optimal plan when $I > 1$.

3.2.2 General case: $j \geq 2$

Solving for the optimal contract involves maximizing the investor's expected payoff with the tools of dynamic programming. As shown in the proof of Proposition 3, this leads to the system of ordinary differential equations recursively defined as:

$$\begin{aligned} (ru + \lambda_j b_j)\, v_j'(u) + j\mu - \lambda_j \left(v_j(u) - v_{j-1}(u - b_j)\right) &= 0, \qquad u \in \left[b_j, \gamma_j\right], & (18) \\ v_j'(u) &= -1, \qquad u \geq \gamma_j, & (19) \end{aligned}$$

where γ_j is a free boundary and $v_j(\cdot)$ is extended into $[0, b_j]$ with the linear interpolation

$$v_j(u) = (u/b_j)\, \overline{v}_j, \qquad \overline{v}_j \equiv v_j(b_j). \tag{20}$$

Proposition 3 *Choose*

$$\frac{r}{\lambda_j} \leq 1 + \frac{\overline{v}_{j-1}}{b_{j-1}} \tag{21}$$

The ordinary differential equation (18)-(19) with linear interpolation (20) has a unique maximal solution v_j for $j \geq 2$. The functions v_j are globally concave and differentiable everywhere except at b_j. The thresholds γ_j

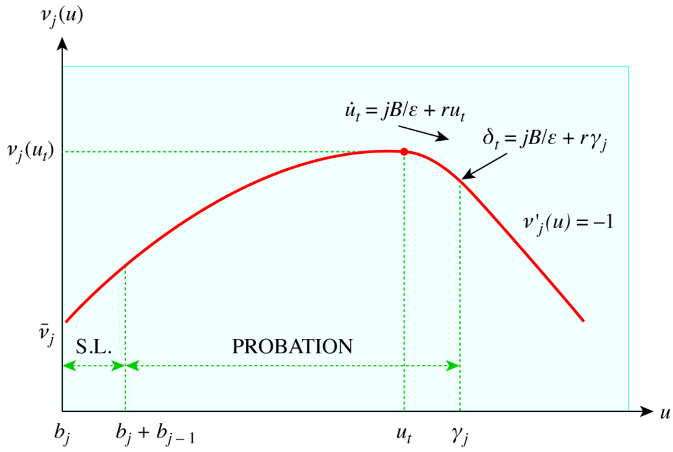

Figure 1: Payoff possibility frontier

are uniquely determined in the intervals $\left[b_j + b_{j-1}, b_j + \gamma_{j-1}\right]$ *by*

$$\frac{r}{\lambda_j} - 1 \in \partial v_{j-1}(\gamma_j - b_j), \tag{22}$$

where $\partial v_j(u)$ *is the subdifferential of* v_j *at* u*. Moreover, if*

$$\frac{r}{\lambda_j} \leq \psi^{-1}\left(\frac{v'_{j-1}(b_{j-1})}{\bar{v}_{j-1}/b_{j-1}}\right), \tag{23}$$

where ψ *is defined by (A.3.1), then* $v'_j(u) \leq v'_{j-1}(u - b_j)$ *for all* $u \in \left[b_j, \gamma_j\right]$*.*

As shown in Figure 1, v_j is strictly concave on $\left[b_j, \gamma_j\right)$. Pool value is maximized at the threshold γ_j, endogenously determined by the boundary condition (19). I call γ_j the "target." The concavity of v_j implies that the shadow price of performance, $1 + v'_j(u)$, is all the more high as performance is low, a property reflecting the inefficiency arising from stochastic liquidation. The highest price arises when utility is at the reservation level b_j. When performance is on target, the shadow price is zero.

Equation (22) is essentially the "smooth pasting" condition $v_j''(\gamma_j) = 0$ ensuring that the target is set optimally. When $\gamma_j > b_j + b_{j-1}$, it simplifies to

$$\frac{r}{\lambda_j} = 1 + v_{j-1}'(\gamma_j - b_j). \tag{24}$$

The interpretation of (24) is the following. The right-hand side is the shadow price of performance right after a default has occurred on target. As long as this is greater than the expected cost of performance, r/λ_j, the target can be raised. Without (21), the backward induction defining γ_j would have to be interrupted at j.

Finally, condition (23) imposes an upper bound on the expected cost of performance r/λ_j. A higher expected cost contributes to raising the shadow price of low performance levels. The shadow price could be so high that it would actually be relieved after a default. This would be inconsistent with optimization. In that case, the optimal plan would have to set a larger penalty than the minimum required for incentive compatibility purposes. The technical condition (23) allows me to abstract from such complications.[21]

3.3 Verification theorem

Proposition 4 *Define u_t as the solution of the stochastic differential equation starting at $u \in [b_I, \gamma_I]$*

$$du_t = (ru_t - \delta_t)\,dt - h_t^1\,(dN_t - \lambda_{I-N_t}\,dt) - h_t^2\,(dH_t - (1-\theta_t)\lambda_{I-N_t}\,dt)\,, \qquad t \in [0, \tau)\,, \tag{25}$$

where on $\{I - N_t = j\}$

$$\begin{cases} \delta_t &= \left(jB/\epsilon + r\gamma_j\right) 1_{\{u_t = \gamma_j\}} \\ \theta_t &= (u_t - b_j)/b_{j-1} \wedge 1 \\ h_t^1 &= (u_t - b_{j-1}) \wedge b_j \\ h_t^2 &= b_{j-1}. \end{cases} \tag{26}$$

Under the conditions of Proposition 3, the optimal contract delivering the bank an expected discounted payoff of u at $t = 0$ given balance I is (δ_t, θ_t) given by (26) and $k = 0$ is incentive-compatible with respect to

[21] Given reasonable parameter values, it was not found stringent for any of the simulations such as the one shown below.

(δ, θ). The processes u_t and $v_j(u_t)$ are the continuation utilities of the bank and the investor, respectively. In particular:

(i) Fees paid to the bank given balance j are $\delta_t = jB/\epsilon + r\gamma_j < j\mu$ if and only if $u_t = \gamma_j$;

(ii) Upon default given balance j, the pool remains in operation with one less unit when $u_t \in \left[b_j + b_{j-1}, \gamma_j\right]$;

(iii) It is maintained with probability $\theta_t = (u_t - b_j)/b_{j-1}$ when $u_t \in [b_j, b_j + b_{j-1})$ and liquidated otherwise.

Investors' strategy relies on two instruments: the promise of future compensation if there is no default for some time (the carrot), and the threat of liquidation if performance is poor (the stick). Consider first the compensation policy. Below target γ_j the strict inequality $v_j'(u) > -1$ indicates that it is cheaper for the investor to compensate the bank with higher continuation utility than to pay cash right away. The rise in continuation utility reflects the prospect of future cash payments. Eventually, the target is reached, unless some default interrupts the process, and the investor prevents u from rising further by paying fees. Compensation δ comprises a management fee equal to a flat percentage of the outstanding pool, jB/ϵ, and a "rent-preserving" component tuned to the bank's discount rate, $r\gamma_j$.

Consider next the liquidation policy. When $u \geq b_j + b_{j-1}$, the penalty incurred by the bank following default in (26) is set at $h^1 = b_j$, the minimum consistent with the incentive-compatible condition (13). There is no threat of liquidation ($\theta = 1$) as the bank's continuation utility after imposition of the penalty, $u - b_j$, is above the reservation level under continuation b_{j-1}. I call $\left[b_j + b_{j-1}, \gamma_j\right)$ the *probation* interval. When $b_j \leq u < b_j + b_{j-1}$, the maximum permissible penalty under (14), $h^1 = u - b_{j-1}$, falls short of the incentive-compatible level b_j because the bank is protected by limited liability. In this event, the optimal response following default is to liquidate the pool with probability $1 - \theta$ in order to restore incentives to monitor even if this is socially costly. The survival probability $\theta = (u - b_j)/b_{j-1}$ reflects the position of u in the stochastic liquidation interval $[b_j, b_j + b_{j-1})$.

The width of the interval $\gamma_j - b_j$ is the maximum buffer protecting the bank against stochastic liquidation and its relative width, $\gamma_j/b_j - 1$, is a measure of how many joint defaults it could withstand starting from target under the current penalty rate. Tuning the target is as effective in disciplining the bank as suspending payments. In normal circumstances — assuming individual risk does not rise after a default — it is not sensible to keep the bank waiting with the promise of larger fees. The reason for actually reducing fees is twofold. First, compensation should not improve in size-adjusted terms when the pool shrinks by one unit. Second, risk-shifting incentives should be held in check and defaults become slightly less frequent, making shirking more difficult to detect. On both accounts, investors' best reaction is to lower the relative size of the buffer, with reservation utility unchanged. This way, a high-performing bank knows that fees will be reduced after a default and keeps monitoring.

But there is a twist. Individual risk may surge if thresholds in the pool structure are reached. Aggregate default intensity can rise despite fewer loans, making it more important to elicit effort at reducing failures. Heightened concerns about credit risk induce a fall in reservation utility, and investors' best reaction is to dampen their impact on the target by raising the relative size of the buffer. By this token, a high-performing bank knows that it has to sustain a relatively long probation period if it operates under turbulence and keeps monitoring.

Although the analysis above rests on simplifying assumptions, the results are qualitatively similar when a few of them are relaxed. In the special case $r = 0$, the bank keeps a stake in the pool at no cost. Accordingly, it always operates on target and the first-best is attained. The pool is never liquidated — investors simply make available to the bank resources that would otherwise be sunk in the pool. A positive liquidation value shrinks the bank's permissible payoff. Intuitively, the bank starts fearing the threat of liquidation at a higher performance level because the pool is worth more in the hands of outsiders when performance is poor. Finally, in the absence of commitment, the bank also fears liquidation at a higher level of performance since it would otherwise try to renegotiate the contract upon default by asking the investor to "forgive" some of the probation period. Details are given in Appendix B.

3.4 Optimality of monitoring

So far, the focus has been on the optimal contract under continuous monitoring. Since $k_t = 0$ is not necessarily efficient when a high rate of impatience r allows for significant gains from trade between the bank and investors,[22] it is important to verify that, given the parameters of the model, it is actually optimal for investors to require a high level of effort from the bank. One has the following result.

Proposition 5 *Under Assumption 2, the optimal contract must involve monitoring at all points in time.*

The intuition is as follows. Given Lemma 2, the bank can only shirk on all or no loans. If the social planner provides the bank with the correct incentives, the bank monitors all loans, and its continuation utility jumps to u_t. This can be sustained in equilibrium only if the rent it extracts, as specified in the efficiency condition (5), is not too large.

3.5 Optimality of stochastic liquidation

The analysis has shown that, even though liquidations are inefficient in the first-best, they are necessary in the second-best in order to restore incentives to monitor when performance is poor.[23] However, liquidation can take many forms. Liquidating all loans with state-dependent probability is not necessarily better than partially liquidating the pool with fixed probability. An option would be downsizing the pool with state-dependent magnitude. Given that liquidations are rarely decided by the flip of a loaded coin, it is important to verify that such liquidation policies cannot improve on social welfare.

Proposition 6 *Suppose that $\bar{v}_j/b_j$ is (weakly) increasing for $j \leq I$. Among all liquidation policies satisfying the incentive compatibility constraint (13) under limited liability, efficiency losses are minimized for the liquidation policy specified in Proposition 4.*

The quantity $\bar{v}_j/b_j$ is the slope of v_j over the interval $[0, b_j]$. The intuition is that, as the balance of the pool increases, investors' value functions v_j expand uniformly upwards. Simulations show that the slope at the origin always steepens when individual default risk remains steady, but subsides if there is a sufficiently

[22] In Plantin (2011), complete securitization with no monitoring can be efficient if gains from trade are sufficiently high, based on a violation of the equivalent of Assumption 2.

[23] In Bolton and Scharfstein (1990), the threat of termination provides the borrower with incentives to repay in staged financing when there is an *ex-post* inefficiency due to the borrower's risk of cash flow diversion. If the setting is modified so that the borrower has all the bargaining power and investors act competitively, the optimal contract of Bolton and Scharfstein also involves stochastic liquidation, with a probability of continuation in the low cash flow state determined by the condition that the incentive compatibility constraint is binding. Stochastic liquidation plays the same incentive role when the inefficiency is due to *ex ante* moral hazard, except that the borrower's reservation utility is defined by a limited liability rather than a truth-telling constraint.

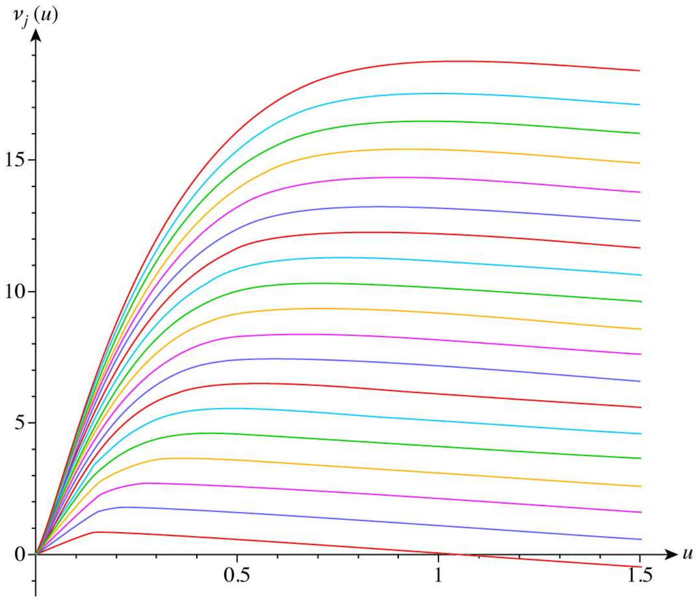

Figure 2: Functions $v_j(u)$ for $j \geq 1$.

large fall in default risk when j increases. Therefore, the qualifying condition of Proposition 6 is more likely to be met if changes in default intensities provided for in Assumption 3 are gradual.[24]

3.6 An example

Given their semi-explicit form, the value functions can be recovered recursively using numerical integration techniques. In this example, we consider a pool of only $I = 20$ unit loans to save on computational costs and take $\mu = 6\%$, $r = 2\%$, $B = 0.2\%$, $\epsilon = 25\%$ as exogenous parameters. Individual default intensities are assumed to be stepwise constant, with $\alpha_j = 5.5\%$ for the first $j \leq 14$ loans informally referred to as the "senior" tranche, $\alpha_j = 5\%$ for the following $15 \leq j \leq 18$ loans or "mezzanine" tranche and $\alpha_j = 4.4\%$ for the last two loans or "junior" tranche $\{19, 20\}$. Assumptions 1, 2 and 3 are satisfied.

The (extended) functions v_j displayed in Figure 2 expand with the size of the pool. Expected discounted payoffs for the bank and investors are maximized for $j = 20$ at $\gamma_{20} = 1.55$ yielding a total value of $v_{20}(\gamma_{20})+$

[24] The capital structure of subprime mortgage-backed securities is typically split up into a large number of tranches (for example 17 in the pool documented by Ashcraft and Schuermann, 2008), so if default intensities are assumed constant over tranches, changes due to the exhaustion of the most junior tranche are more likely to be gradual.

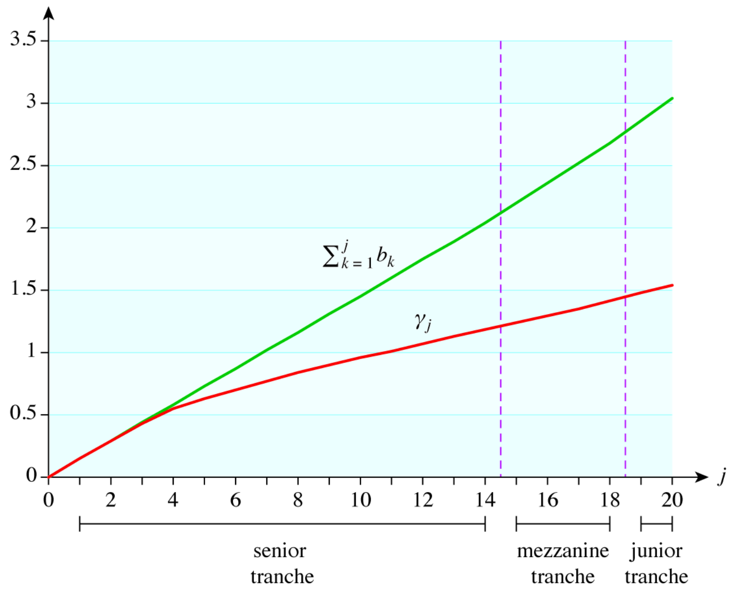

Figure 3: Economic interest vs. risk requirements

$\gamma_{20} = 21.14$, in between the initial investment and the social optimum $I\mu/\bar{\alpha}_I = 22.8$ implied by the first-best surplus (10). The reduction in value due to the agency problem is above 6%, a loss of about 60% in terms of social surplus. At the cut-off date investors will lend no more than $v_{20}(\gamma_{20}) = 19.59$, implying that the bank should contribute 0.41, more than 2%, from its own funds.

An important feature implied by Proposition 3 is that the target cannot increase by more than the penalty rate, as $\gamma_j \leq b_j + \gamma_{j-1}$. Hence γ_j is less than the sum of penalties $\sum_{k=1}^{j} b_k$ incurred down the road if the pool is monitored until extinction. The former variable is the bank's continuation utility under target, i.e. the highest economic interest it can afford given the current size. The latter variable is the risk it is required to retain over time on order to maintain the credit quality of the outstanding pool. Because risk retention is costly when $r > 0$, the rent extracted by the bank at any point in time cannot be as high as the risk it is required to retain. In particular at time zero, the bank's economic interest can be no more than $\gamma_{20} = 1.55$, while the risk retained is $\sum_{j=1}^{20} b_j = 3.04$, almost twice as large. This is in sharp contrast with actual securitization practice, where economic interest refers to the amount of risk retention rather than the value of future discounted cash flows.

The targets γ_j as well as the risk retention requirements $\sum_{k=1}^{j} b_k$ displayed in Figure 3 increase with the size of the pool. Risk requirements are higher in the mezzanine tranche and highest in the junior tranche, as the bank needs high-powered incentives when default intensity is low. This causes the slope of the risk retention curve to steepen with more junior tranches. On the other hand, the bank's economic interest grows at a lower pace, indicating that the reduction in fees (when paid) is lower after defaults accumulate. When the "mezzanine" and "senior" tranches are reached, reservation utility b_j falls by 12% and 9% respectively, due to the higher credit risk involved. The target γ_j falls only by 4.6% and 4.9% respectively, implying that the range $\left[b_j, \gamma_j\right]$ is wider in proportion.

4 Implementation

Given the history of losses, fees can be reverse-engineered from u_t using the optimal controls (26). A natural way of implementing the compensation policy is to replicate u_t dynamically by use of a *cash reserve account* that faithfully tracks the implications of losses for the bank's continuation utility. The account is managed by the trust and actually plays two roles. One is to provide protection to investors. The other is to govern performance-based compensation. The current balance reveals outright performance and can be used to determine the amount and timing of fees that are released.

4.1 Capital required

With competitive investors, the bank maximizes profits at time 0 as

$$\begin{aligned} \max_{\{u,K\}} \quad & u - K \\ \text{s.t.} \quad K \ & \geq \ I - v_I(u), \end{aligned} \tag{27}$$

which yields $u \geq \gamma_I$ and $K = I - v_I(u)$. The bank's profit is the social surplus $S = u + v_I(u) - I$. It is positive for a "premium" pool — a pool of loans whose securitization can generate deemed proceeds in excess of the principal amount — and constant over the linear portion of investors' value function. Irrespective of the value initially chosen for u, the amount lent by investors is always the pledgeable part $v^* = v_I(\gamma_I)$ and the bank starts with continuation utility $u^* = \gamma_I$. The bank allocates internal capital to the pool in the amount $K = I - v^*$.

The optimal contract can be implemented through securitization in the following sense. The bank first initiates an ABS transaction by selling the pool to a bankruptcy-remote trust with gain on sale S over principal balance I. In a typical *partial* securitization, the bank would retain K as "skin in the game" in the form of ABS interests. However, the premium S would qualify under generally accepted accounting principles (GAAP) as gain on sale only if the bank does not have a "controlling financial interest" in the trust and the transferred assets remain in "legal isolation" even after bankruptcy or receivership. I defer the discussion of such practices until Section 5 and describe instead the implementation of a *complete* securitization scheme.

The tranching can be done as follows. Tranche $k = 1, \ldots, K$, with lower attachment point L^k and upper attachement point U^k, has total outstanding $U^k - L^k$, where $0 = L^1 < \cdots < U^k = L^{k+1} < \cdots < U^K = I$. The cumulated percent width of all junior tranches, L^k/I, is referred to as subordination. Once the pool size j hits $I - L_k$, there is no more subordination and losses are applied to the tranche. Assume that individual intensities α_j are stepwise constant. In that case, the senior-subordinated structure can be adjusted so that the attachment points match the jumps in α_j. This implies that the penalties $b_j = B/(\epsilon\alpha_j)$ are a fixed percentage, noted b^k, of the writedown amounts as long as $N_t = I - j \in \left[L^k, U^k\right)$.

The next proposition suggests that risk retention alone cannot correct misaligned incentives, but that liquidity regulation can bring them back to the fold.

Proposition 7 *With constant individual risk within tranches, optimal ABS can be implemented as follows:*

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

(ii) The sponsor sells protection via ABCDS of notional $b^k(U_k - L_k)$*, where* $b^k = b_j$ *if* $I - j \in [L_k, U_k)$*;*

(iii) The protection imbedded in ABCDS is assigned to the reserve account;

(iv) The management fee Bj/ϵ *and accrued interests* (ru_t) *are applied to the reserve account;*

(v) The account balance is maintained between cap γ_j *and floor* b_j*:*

— Excess cash triggers payment to the bank;

— *Overdrafts trigger stochastic liquidation: the trust makes up for the shortfall if the pool is maintained, or seizes the account if the pool is liquidated, with continuation probability* $\theta_t = (u_{t-} - b_j)/b_{j-1}$ *equal to the balance following the overdraft in percentage of the new floor.*

4.2 Skin in the game

With a complete securitization scheme, the bank is no longer entitled to the ABS interests issued by the trust. Its economic exposure takes the form of a credit enhancement. It returns $u^* = K + S$ in a cash reserve account, which is used as collateral for ABS credit default swaps written to reflect how much risk it should retain over time.

In contrast to a generic CDS where failure to pay is a one-off event, ABCDS are "pay-as-you-go" contracts committing the protection seller to reimburse writedowns on a single ABS tranche. The notional of the contract declines over time with principal writedowns. Given an "applicable percentage" of b^k — the ratio of the CDS notional to the outstanding of the ABS — ABCDS k reimburses a fraction b^k of writedowns between L^k and U^k as they occur.

Defaults act as a trigger mechanism to make "margin calls" in the amount of the bank's risk retention shares $b_j = b^k$ when ABS k is the most junior tranche remaining. In a margin account, the broker asks the investor to post new collateral as prices drop. Here, the protection is posted in advance through ABCDS contracts and covered by withdrawals from the reserve account.[25] A portfolio of ABCDS with appropriate applicable percentages ensures that risk retention as implied by the optimal contract is enforced. "Retaining" the risk means that the bank writes protection to the trust. It does not enter into a contract with itself but simply owns the risk by not receiving the premium which would normally be paid by the trust as the buyer of protection.

Assumption 3 implies that the bank's risk retention shares taper off as losses unfold. In the beginning, the bank needs high-powered incentives and bears the brunt of initial losses. In the end, underlying risk is

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

high relative to private benefits and the bank is better shielded against losses.[26] Note that since ABCDS are "bets" on a pool making losses in specific tranches, their prices can be used to extract the underlying default intensities as perceived by investors at the time of issuance and infer the requisite risk retention amounts.

4.3 Cash reserve management

Beyond providing a cushion against losses, the cash reserve account also provides a mechanism for regulating compensation. The idea is that if the account balance reflects the bank's stake as in the optimal contract, its movements faithfully mirror pool performance over time. It is maintained between a size-dependent cap γ_j (target level) and a tranche-dependent floor b_j (reservation level). The role of the trust is to capture "unused" bank fees in the amount $\delta'_t = jB/\epsilon + ru_t$ for potential future use when $u_t < \gamma_j$. In this event, management fees and accrued interests are applied to the reserve account rather than paid. If bank fees have built up the balance to the cap, they become available for distribution. If the protection written causes the balance to drop after a default, fees collected are again applied to the reserve account until the balance is restored to the new cap.

The reserve account provides protection against losses that do not generate an overdraft. When an ABCDS writedown charged against it depletes the balance to a level which would effectively negate the bank's incentives, i.e. when $u_{t-} - b_{I-N_{t-}} < b_{I-N_{t-}-1}$, the trust is responsible for stochastic liquidation. This can be achieved by a *clean-up call* option.[27] It gives the trust the right to repurchase the pool at the (zero) liquidation price when the balance falls below floor. The trust's cumulated cost resulting from intervention during stochastic liquidation episodes is given by the martingale

$$\xi_t = \int_0^t \left[b_{I-N_{s-}} + b_{I-N_{s-}-1} - u_{s-}\right]^+ dN_s - \int_0^t b_{I-N_{s-}-1}\, dH_s. \tag{28}$$

With probability θ_t, the call option is not exercised and the trust settles the shortfall with a cash payment

[26] The interplay between deferred compensation and risk retention is also analyzed in a screening context by Inderst and Pfeil (2010), where both internal and external agency problems between the bank, its loan servicers and outside investors are modeled.

[27] Most ABS deals have clean-up call features which can be triggered if the pool balance falls to 10% of the original amount.

$\Delta\xi_t = b_{I-N_{t-}-1} - \left(u_{t-} - b_{I-N_{t-}}\right)$ to keep the balance afloat. With probability $1-\theta_t$, the trust seizes the account, wins the residual balance $-\Delta\xi = u_{t-} - b_{I-N_{t-}}$, and divests the bank and investors of any rights to the pool, which is liquidated.

Under Proposition 7, the trust breaks even. The arrangement is consistent with separating different functions, with the bank in charge of the continuous monitoring on the one hand, and investors related to the securitization through their ABS interests on the other. The sponsoring bank maximizes profits subject to its conducting due diligence. Thus, investors can price their ABS interests at the cut-off date by taking into account any credit enhancement structure that they see fit and the risk of stochastic liquidation.

5 Comparison with proposed rulemaking

The unprecedented tide of foreclosures that swept over the subprime market in the summer of 2007 led policy makers to focus their attention on the senior-subordinate structures that were used in home equity ABS. Subordination levels were in most cases insufficient to absorb the overwhelming losses made on their collateral, triggering rating actions or writedowns on even the most senior securities. A few of the key features of the NPR, as well as the comments made by financial institutions acting as securitization sponsors, can be usefully highlighted from the vantage point of the model.

5.1 Premium capture

A novel feature in the NPR that has attracted widespread criticism from ABS sponsors is found in the "premium capture" provisions. In the Agencies' view, sponsors were able to reduce or even negate the economic interest they had retained in securitization transactions by selling premium or interest-only tranches. Because this creates incentives for poor underwriting, sponsors would be prevented from extracting value at closing by monetizing the "excess spread" — the excess of the net interest rate earned on the pool over the weighted-average coupon paid on the related securities — to which holders of the residual interest are entitled. This would be achieved by requiring them to place the transaction premium in a Premium Capture

Cash Reserve Account (PCCRA) at issuance. The amount would be equal[28] to the excess of all proceeds received from the sale of ABS interests to persons other than the retaining sponsor $(I + S - K)$ over 95% of par value of the ABS issued (95% I), or in terms of the model $S + 5\%\ I - K$, where $K \geq 5\%\ I$ is the amount of risk retention excluding qualified residential mortgages under the Dodd-Frank Act. Cash deposited in the PCCRA would be invested in low-yielding investments, subordinated to any other ABS interest in the securitization and maintained for the life of the transaction.

The premium capture provisions effectively subordinate the gain on sale arising from the securitization of a premium pool. The sponsoring bank can either retain risk as a residual (or other ABS) interest in the amount $K = 5\%\ I$ and establish a premium capture account S exposed to losses on the pool even before the residual interest. Or else it can reduce the funding of the premium capture account but then is forced to increase the par value of the residual (or other ABS) interest by the same amount. Despite the pleas for withdrawal made by the mortgage side of the ABS industry, based in part on a moot discussion of the sense of "par value" in the NPR, the model suggests that the Agencies' view is correct. The economic interest retained by the sponsor at inception is $u^* = K + S$ under Proposition 7, indicating that subordination of the premium must be accomplished *in addition to* that of the capital charge that the bank has to hold against the pool exposures. It is interesting to note that, on a theoretical basis, subordination of the premium is actually *implied* by the profitability of securitizations.

One reason the PCCRA appears cost-prohibitive to the industry is not so much that it traps funds in addition to the five percent risk retention requirement. Rather, it locks them into a limited range of investments whose yields are not commensurate with their subordination level. The NPR provides that cash reserves can substitute for risk retention if they are fully funded at issuance, but subordinating the PCCRA to the first-loss position within the securitization, without proper compensation, makes the option uneconomical. To this extent, the PCCRA might be amended to allow uniform subordination across the premium and the capital charge. According to the model, this could be done by using the PCCRA as collateral for unfunded risk retention, rather than a layer of protection grafted onto an existing funded

[28] This is assuming away the "representative sample" form of retention under which only part of the pool is securitized.

ABS-interest scheme.

5.2 Risk retention

Credit risk retention rules under the NPR are intended to be flexible in their implementation to accommodate a large heterogeneity across asset classes. As mentionned in the Introduction, sponsors have various options in the slicing of their ABS interests, such as the "vertical" form of risk retention (5% of each ABS class issued), the "horizontal" one (first-loss position in an amount of 5% of all ABS issued) or the "L-shaped" one (a combination of the two). In the model, in contrast, sponsors are required to share in the risk at a declining rate until liquidation, as risk retention is linked to the need to restore incentives when monitoring is not observable.

The proposed risk retention rules are considered very blunt by the securitization industry, which argues for more flexibility in their implementation. The model's normative content in this respect is weak as the risk retention shares obtained are driven by specific assumptions about banks' incentives and contagion risk.[29] However, the merits of flexibility should be balanced against the concern that intermediaries' retention might be inconsistent with the influence they retain through their monitoring. Their choices have different implications for capital requirements under Basel II and may be driven by regulatory arbitrage. The paper concurs with Franke and Krahnen (2009) that the allocation of risk is probably quite different from what theory predicts and remains of particular relevance for bank supervisors.

More importantly, the securitization industry has argued that, insofar as risk retention mechanisms are designed to prevent the poor origination of credit, they should be removed after a few years, when losses are driven by factors beyond the reach of originators. However, some provisions in the NPR indicate that regulators may be uncomfortable with that view. For example, the inclusion of mandatory servicing standards in the definition of qualified residential mortgages for risk retention exemption suggests that lenders have obligations vis-à-vis borrowers beyond lending them money and that poor servicing may result in excess delinquencies many years into the origination. A reserve account could help bridge this gap because

[29] For example, if banks had to appraise the quality of borrowers before marketing the loans, a disproportionate share of the risk would be allocated to the bank ($b_j = 1$) to provide it with an equity position in the loans; see Pennachi (1988).

it would "delink" risk retention from economic interest and enable sponsors to extend protection in a much more cost-effective way than currently permitted under the NPR.

Enlarging the role of the reserve account from only capturing the premium to collateralizing risk retention would have several advantages. First, withdrawals from the reserve account when ABCDS payments are made contribute to lowering the account balance on average, thus saving on costly resources when sponsors' rates of return are high. Second, fees withheld when performance is poor can replenish the reserve account and bolster its role as a cushion against future losses, allowing total risk exposure to far exceed the current balance. Finally, the unhedged part of the risk exposure is actually borne by the trust due to the sponsor's limited liability in the event of a liquidation. The ABCDS premium flows waived by the sponsoring bank can be viewed as a prepayment for the contingent support it receives, in the spirit of the capital insurance scheme proposed by Kashyap, Rajan and Stein (2008). When losses begin unfolding, capital is automatically supplied by the sponsoring bank from the cash reserve account and the tax is high. Only under liquidation is the capital overwhelmingly supplied by the trust and the liquidity tax eventually eschewed.

5.3 Eligible horizontal interest

In a typical home equity loan securitization, excess spread and overcollateralization (OC) — the mismatch between the pool balance and the balance of coupon-bearing ABS securities — are critical components of the deal's credit enhancement. Current period losses are first allocated to the excess spread. The "unused" portion of excess spread, if any, is then applied to pay down the principal balance of senior tranches. This *accelerated amortization* increases the OC[30] until the subordination of the senior class (the percentage width of the subordinate classes, including the OC) is twice as large as the original level. The point at which express spread has built up senior subordination to the target level is known as the *stepdown date*, and must occur after a *lockout period* of generally three years. After the stepdown date, subordinate and residual classes become eligible to receive distributions of principal.

The OC provides a cushion against future losses, much like the reserve account in the model. Any excess

[30] The OC also increases in percentage following the distributions of principal that are only applied to the senior class over that period.

OC over its own target level (usually twice the OC as of the cut-off date) is released to the residual holder, provided certain performance tests[31] are met. If current period losses exceed excess spread, they are allocated to the OC and excess spread collected in later months is not released but applied to amortize outstanding bonds until the OC is restored to its target level. Hence the OC functions as the most subordinated class contained within the residual interest. Accordingly, the NPR defines an " eligible horizontal residual interest" as one that has the most subordinated claim on cash flows and is allocated all losses not covered out of excess spread until reduced to zero.

Despite the similarities, there is an important difference. Under horizontal risk retention, the sponsor has an equity position and is entitled pro rata to all excess spread released to the residual interest. If losses occur late in the life of the deal and prepayments are slow, excess spread is abundant and the residual holder absorbs a disproportionate share of the premium. As shown by the premium capture provisions, extracting value from the excess spread creates incentives to maximize the securitization scale and is inconsistent with the optimal plan. This is true whether the claim to excess spread is sold to investors at inception or captured by the retained residual over the life of the deal. In contrast, the reserve account contributes no more than the payment of interest established at closing, much like the PCCRA, were it not for the fact that no replenishment of the PCCRA is expected from cash flow diversion when performance is poor.

5.4 Consolidation versus sale

Finally, the premium capture provisions have been criticized for their interaction with GAAP. As mentionned in Section 4, a transaction qualifies for derecognition only if it satisfies consolidation and sales accounting standards. The Board of Governors of the Federal Reserve (2010) has expressed the view that horizontal risk retention exposes the holder to benefits and losses that "could potentially be significant" and is more likely to result in the sponsor being viewed as having a controlling financial interest in the trust. The industry's claim is that the funding of a premium account, which covers losses before they are allocated to reduce the principal of any ABS interest, is akin to imposing an additional horizontal risk retention on the sponsor.

[31] OC structures generally provide two trigger tests, one based on delinquency rates and the other on cumulative losses.

Under this interpretation, the transaction is likely to be consolidated instead of achieving off-balance sheet accounting treatment.

Although the PCCRA would increase the amount of risk retained and make it more difficult to deconsolidate transferred assets, the use of a reserve account as collateral for risk retention would preclude all uncertainty as to whether a true sale can be achieved. The thrust of the model is that sponsors are required to retain risk as the solution to a principal-agent problem where the power to direct activities is held by investors and the obligation to absorb losses is dictated by optimal controls that are beyond the reach of the agent. Under these circumstances, it is doubtful that its implementation could be viewed as providing the sponsor with a controlling financial interest in the trust.

6 Conclusion

Under the optimal mechanism, the continuation utility of a sponsoring bank decreases after a shock. The fall is related to the length of the probation period during which the bank receives no payment and improves its performance record. The payment received after probation gets smaller as the number of shocks grows, and liquidation is possible when performance is sufficiently poor.

Pool performance is difficult to measure. As pointed out by Tirole (2011), an important regulatory issue is whether dynamic management of liquidity has a role to play beyond the usual solvency requirement. The upshot of the model is that, by properly managing liquidity, the trust can reveal the market value of pool performance and remove all uncertainty about the underlying quality of monitoring, thus obviating the need to rely on a proprietary quantitative risk model or some expert judgement exercised by bank supervisors. Drawing down the liquidity position does not leave the buffer exposed to the "repeated" liquidity shock conundrum (Goodhart, 2008).

While the literature generally considers endogenous liquidation values with exogenously given contracts (Schleifer and Vishny, 1992), here I endogenize contracts with exogenously given liquidation values. I find under crude assumptions that bank pay, capital requirements, securitization and liquidity risks should not be regulated separately, each under its own "global framework."

As usual in the incentives-based corporate finance literature, the paper has drawn implications for the capital structure of the bank, even though it is strictly about the compensation of the loan manager. These are in principle unrelated things. A better account of moral hazard in subprime morgage securitizations would be warranted, especially regarding servicers' forbearance or short-termism (Edmans et al., 2011) in the event of delinquencies. The model could also be extended in a number of interesting directions. Most toxic mortgages before the crisis were in the form of interest-only or balloon mortgages, with low payments and default risk in the beginning, and high payments and default risk in the end. Other, more realistic specifications of risks would also make sense, such as good or bad luck (Hoffmann and Pfeil, 2010) caused by news about home prices or shifts in monetary policy. How this would modify the optimal contract is left for future research.

7 References

Adelino, M., Gerardi, K., Willen, P., 2009. Why don't lenders renegotiate more home mortgages? Redefaults, self-cures and securitization. NBER Working Paper 15159.

Agarwal, S., Lucca, D., Seru, A., Trebbi, F., 2012. Inconsistent regulators: Evidence from banking. NBER Working Paper 17736.

Aït-Sahalia, Y., Cacho-Diaz, J., Laeven, R., 2010. Modeling financial contagion using mutually exciting jump processes. Working Paper, 2010.

Ashcraft, A., Schuermann, T., 2008. Understanding the securitization of subprime mortgage credit. Federal Reserve Bank of New York Staff Reports, 318.

Azizpour, S., Giesecke, K., Kim, B., 2011. Premia for correlated default risk. Journal of Economic Dynamics and Control, 35, 1340–1357.

Besanko, D., Kanatas, G., 1993. Credit market equilibrium with bank monitoring and moral hazard. Review of Financial Studies, 6(1), 213–232.

Biais, B., Mariotti, T., Plantin, G., Rochet, J.-C., 2007. Dynamic security design: Convergence to continuous time and asset pricing implications. Review of Economic Studies, 74, 345–390.

Biais, B., Mariotti, T., Rochet, J.-C., Villeneuve, S., 2010. Large risks, limited liability and dynamic moral hazard. Econometrica, 78(1), 73–118.

Board of Governors of the Federal Reserve, 2010. Report to Congress on Risk Retention. Available at http://federalreserve.gov/boarddocs/rptcongress/securitization/riskretention.pdf.

Bolton, P., Scharfstein, D., 1990. A theory of predation based on agency problems in financial contracting. American Economic Review, 80(1), 93–106.

Cantor, R., Hu, J., 2007. Deal sponsor and credit risk of asset-backed and mortgage-backed securities. Available at SSRN: http://ssrn.com/abstract=996014.

Davis, M., Lo, V., 2001. Infectious defaults. Quantitative Finance, 1, 382–387.

Dell'Ariccia, G., Igan, D., Laeven, L., 2009. Credit booms and lending standards: Evidence from the subprime mortgage market. Tilburg University Discussion Paper 46.

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

DeMarzo, P., Fishman, M., 2007a. Agency and optimal investment dynamics. The Review of Financial Studies, 20, 151–189.

DeMarzo, P., Fishman, M., 2007b. Optimal long-term financial contracting. The Review of Financial Studies, 20, 2079–2128.

Demyanyk, Y., van Hemert, O., 2011. Understanding the subprime mortgage crisis. Review of Financial Studies, 24(6), 1848–80.

Diamond, D., 1984. Financial intermediation and delegated monitoring. Review of Economic Studies, 51, 393–414.

Dichev, I., Skinner, D., 2002. Large-sample evidence on the debt covenant hypothesis. Journal of Accounting Research, 40(4), 1091–1123.

Duffie, D., 2008. Innovations in credit risk transfer: Implications for financial stability. BIS Working Paper 255.

Edmans, A., Gabaix, X., Sadzik, T. and Sannikov, Y., forthcoming. Dynamic CEO compensation. Journal of Finance.

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

Fender, I., Mitchell, J., 2009. Incentives and tranche retention in securitisation: A screening model. BIS Working Paper 289.

Financial Services Roundtable, Comments of Housing Policy Council, 2011. Credit risk retention. Available at http://www.fsround.org/fsr/policy_issues/regulatory/pdfs/pdfs11/riskretentionletterQRM.pdf.

Fitch, 2003. Residential mortgage servicer ratings. Dated February 21.

Franke, G., Krahnen, J., 2006. Default risk sharing between banks and markets: the contribution of collateralized debt obligations, in: The Risks of Financial Institutions, NBER Conference Report. University of Chicago Press, Chicago and London, 603–631.

Franke, G., Krahnen, J., 2009. The future of securitization, in: Fuchita, Y., Herring, J., Litan, E. (Eds.), Prudent Lending Restored: Securitization after the Mortgage Meltdown. Brookings Institution Press, Washington, D.C., 105–61.

Frey, R., Backhaus, J., 2008. Pricing and hedging of portfolio credit derivatives with interacting default intensities. International Journal of Theoretical and Applied Finance, 11(6), 611–34.

Gan, Y., Mayer, C., 2006. Agency conflicts, asset substitution, and securitization. NBER Working Paper 12359.

Giesecke, K., Spiliopoulos, K., Sowers, R. and Sirignano, J., 2011. Large portfolio asymptotics for loss from default. Mimeo.

Greenbaum, S., Thakor, A., 1987. Bank funding modes: Securitization versus deposits. Journal of Banking and Finance, 11, 379–401.

Goodhart, C., 2008. Liquidity risk management. Banque de France Financial Stability Review, 11, 39–44.

Gorton, G., Pennacchi, G., 1995. Banks and loan sales: Marketing nonmarketable assets. Journal of Monetary Economics, 35, 389–411.

Hart, O., Zingales, L., 2008. A new capital regulation for large financial institutions. CEPR Discussion Paper 7298.

Hartman-Glaser, B., Piskorski, T., Tchistyi, A., 2012. Optimal securitization with moral hazard. Journal of Financial Economics, 104, 186–202.

Hoffmann, F. and Pfeil, S., 2010. Reward for luck in a dynamic agency model. Review of Financial Studies, 23(9), 3329–45.

Holmström, B., Tirole, J., 1997. Financial intermediation, loanable funds, and the real sector. Quartely Journal of Economics, 112(3), 663–691.

Inderst, R., Pfeil, S., 2010. Securitization and compensation in financial institutions. CEPR Discussion Paper 8089.

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

International Monetary Fund, 2008. Containing Systemic Risks and Restoring Financial Soundness, Global Financial Stability Report. Available at http://www.imf.org/External/Pubs/FT/GFSR/2008/02/index.htm.

Jarrow, R., Yu, F., 2001. Counterparty risk and the pricing of defaultable securities. Journal of Finance, 53, 2225–2243.

Kashyap, A., Rajan, R., Stein, J., 2008. Rethinking capital regulation, in: Federal Reserve Bank of Kansas City Symposium at Jackson Hole. Available at http://www.kc.frb.org/publicat/sympos/2008/KashyapRajanStein.03.12.09.pdf.

Keys, B., Mukherjee, T., Seru, A., Vig, V., 2010. Did securitization lead to lax screening? Evidence from subprime loans. Quartely Journal of Economics, 125(1), 307–62.

Laurent, J.-P., Cousin, A., Fermanian, J.-D., 2008. Hedging default risks of CDOs in Markovian contagion models. Available at DefaultRisk: http://www.defaultrisk.com/pp_cdo_26.htm.

Moody's, 1999. Special servicing and default management in the subprime mortgage market: The loan doctors. Special Report.

Norden, L., Weber, M., 2010. Credit line usage, checking account activity, and default risk of bank borrowers. Review of Financial Studies, 23(10), 3665–99.

O'Keefe, J., 2010. The effects of underwriting practices on loan losses: Evidence from the FDIC survey of bank lending practices, in: Kim, S.-J., Mckenzie, M. (Eds.), International Banking in the New Era: Post-Crisis Challenges and Opportunities. International Finance Review, 11, Emerald Group Publishing Limited, pp. 273–314.

Pagès, H., Possamaï, D., 2012. A mathematical treatment of bank monitoring incentives. Banque de France Working Paper 378.

Pennacchi, G., 1988. Loan sales and the cost of bank capital. The Journal of Finance, XLIII(2), 375–396.

Pennington-Cross, A., Ho, G., 2006. Loan servicer heterogeneity and the termination of subprime mortgages. Federal Reserve Bank of St. Louis, Working Paper 24.

Piskorski, T., Seru, A., Vig, V., 2010. Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis. Journal of Financial Economics, 97(3), 369–97.

Plantin, G., 2011. Good securitization, bad securitization. Institute for Monetary and Economic Studies, Bank of Japan Discussion Paper 2011-E-4.

Sannikov, Y., 2008. A continuous-time version of the principal-agent problem. Review of Economic Studies, 75, 957–984.

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

Securities Industry and Financial Markets Association, 2011. Credit Risk Retention: Proposed Rule. Available at http://www.sifma.org/issues/item.aspx?id=24954.

Tirole, J., 2011. Illiquidity and all its friends. Journal of Economic Literature, 49(2), 287–325.

Yu, F., 2007. Correlated defaults in intensity-based models. Mathematical finance, 17(2), 155–173.

A Sketches of Proofs

A.1 Lemma 1

The bank's expected *lifetime utility*, conditional on $\mathcal{G}_t$, is

$$\begin{aligned} U_t^k &= E^k \left[\int_0^\tau e^{-rs} \left(\delta_s + B k_s \, ds \right) \Big| \, \mathcal{G}_t \right] \\ &= \int_0^{t \wedge \tau} e^{-rs} \left(\delta_s + B k_s \right) ds + e^{-rt} u_t^k. \end{aligned} \tag{A.1.1}$$

Since U_t^k is a $\mathcal{G}_t$-martingale under P^k, the martingale representation theorem for point processes implies that there are predictable processes h^1 and h^2 such that

$$dU_t^k = -e^{-rt} h_t^1 \left(dN_t - \lambda_t^k \, dt \right) - e^{-rt} h_t^2 \left(dH_t - (1 - \theta_t) \lambda_t^k \, dt \right) \tag{A.1.2}$$

on $\{t \leq \tau\}$. Differentiating (A.1.1) and substituting in (A.1.2) yields (12).

A.2 Lemma 2

Consider an arbitrary strategy $\widehat{k} \neq 0$ specifying the number of unmonitored loans at any point in time until liquidation. Let u_t denote the continuation utility in (8) resulting from the all-out effort strategy $k = 0$. Define by

$$\widehat{U}_t = \int_0^{t \wedge \tau} e^{-rs} \left(\delta_s + B \widehat{k}_s \right) ds + e^{-rt} u_t$$

the lifetime utility of the bank viewed as of time t if it follows the strategy $\widehat{k}$ before time t, and plans to switch to $k = 0$ afterwards. One can show that, under $P^{\hat{k}}$, the drift of $\widehat{U}$ is $e^{-rt} \left(B - \epsilon \alpha_{I - N_{t-}} (h_t^1 + (1 - \theta_t) h_t^2) \right) \widehat{k}_t$. Then we know the strategy $k = 0$ is at least as good as $\widehat{k}$ if and only if this drift is less than or equal to zero almost everywhere. The proof follows along exactly the same lines as in the proof of Proposition 2 of Sannikov (2008). This yields (13), as desired.

A.3 Proposition 3

Under the incentive compatibility constraint (13), the bank chooses $k = 0$ and (12) takes the form of the ordinary differential equation

$$\dot{u}_t + \delta_t = ru_t + \lambda_j \left(h_t^1 + (1-\theta_t)h_t^2\right)$$

on the stochastic time interval $\{I - N_t = j\}$. This leads to the following system of Hamilton Jacobi Bellman (HJB) equations to be solved recursively from $j = 1$ to I

$$\begin{aligned} \max_{\{\delta\geq 0,\theta\in[0,1],h^1,h^2\}} \quad & \left\{\left(ru + \lambda_j(h^1 + (1-\theta)h^2) - \delta\right) v_j'(u) + j\mu - \delta \right. \qquad (\mathcal{P}) \\ & \left. -\theta\lambda_j \left(v_J(u) - v_{j-1}(u-h^1)\right) - (1-\theta)\lambda_j v_j(u)\right\} = 0, \\ & \text{subject to (13), (14) and (15),} \end{aligned}$$

with initial condition $v_0(u) = 0$. Optimizing with respect to the controls, one finds (26) if $v_j'(u) \leq v_{j-1}'(u-b_j)$ for all $u \in \left[b_j, \gamma_j\right)$, where γ_j is the smallest u such that $v_j'(u) = -1$. Substituting into $(\mathcal{P})$, one finds (18)-(19) with linear interpolation (20).

We first show by induction that for any $\gamma_j \in \left[b_j + b_{j-1}, b_j + \gamma_{j-1}\right]$ the extended function v_j is differentiable everywhere except at b_j and concave on $\left[b_j, \gamma_j\right]$. By evaluating (18) at b_j we have

$$\frac{\overline{v}_j}{b_j} = \frac{\mu\epsilon}{B} + \frac{r+\lambda_j}{\lambda_j} v_j'(b_j),$$

implying $\overline{v}_j/b_j - v_j'(b_j) \geq \mu\epsilon/B - r/\lambda_j$, which can be shown to be positive under Assumption 2. Thus, the extended function v_j is also concave on $\left[0, \gamma_j\right]$.

Second, the function v_j is maximal if the smooth-pasting condition $0 \in \partial v_j'(\gamma_j)$ is satisfied. Differentiating (18) and evaluating at $u = \gamma_j$ yields (22). As shown on Figure 4, a solution exists if and only if (21) is met. Either r/λ_j is in the interval $\left[1 + v_{j-1}'(b_{j-1}), 1 + \overline{v}_{j-1}/b_{j-1}\right]$ and $\gamma_j - b_j = b_{j-1}$ or $r/\lambda_j < 1 + v_{j-1}'(b_{j-1})$ and $b_{j-1} < \gamma_j - b_j \leq \gamma_{j-1}$. In either case, the extended value function v_j is concave over $[0, \infty)$. Note that $\gamma_j - b_j \leq \gamma_{j-1}$ implies that $\delta_t \leq jB/\epsilon + r\gamma_j \leq jB/\epsilon\,(1 + r/\bar{\alpha}_j) \leq j\mu$ under Assumptions (2) and (3).

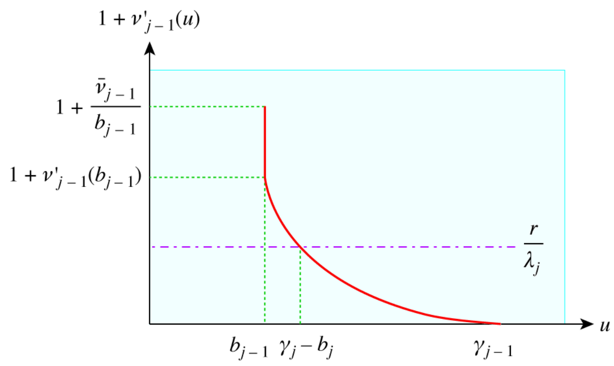

Figure 4: Smooth-pasting condition

Finally, we find a sufficient condition under which $v_j'(u) \leq v_{j-1}'(u - b_j)$ for all $u \in [b_j, \gamma_j]$. The condition is technical; cf. Pagès and Possamaï (2011) for details. The fonction ψ in Proposition 3 is defined as

$$\begin{aligned} \psi(x) &\equiv \frac{\phi(x) - x}{(1-x)\,\phi(x)}, \qquad x > 0 \\ \phi(x) &\equiv \left(\frac{1+x}{1+2x}\right)^{1/x-1}. \end{aligned} \tag{A.3.1}$$

It is decreasing from 1 to 1/2 and can be inverted with the convention $\psi^{-1} = \infty$ on $[0, 1/2]$.

A.4 Proposition 4

See Pagès and Possamaï (2011).

A.5 Proposition 5

Under the optimal policy, pool value $u_t + v_j(u_t)$ is governed by Proposition 4. Suppose instead that the social planner allows the bank to reap private benefits over $[t, t+dt)$ and reverts to the optimal policy afterwards. Underlying risk jumps from λ_j to $(1+\epsilon)\,\lambda_j$ during the infinitesimal time interval. On the other hand, the social planner can dispense with the promise-keeping constraint[32] as the bank fails to monitor anyway.

[32] Including interest rate charges in the dynamics of pool value under the alternative policy would only reinforce the conclusion.

Let W_{t+dt} be the pool value associated with the optimal policy at date $t+dt$ and $j = I - N_t$. If $u_t \in \left[b_j + b_{j-1}, \gamma_j\right]$, the contract is in state (j, u_t) if $dN_t = 0$ or in state $(j-1, u_t - b_j)$ if $dN_t = 1$. The social loss incurred in case of default is

$$\Delta W_{t+dt} = v_j(u_t) - v_{j-1}(u_t - b_j) + b_j. \tag{29}$$

If $u_t \in [b_j, b_j + b_{j-1})$, the contract is in state (j, u_t) if $dN_t = 0$, in state $(j-1, b_{j-1})$ if $dN_t = 1$ and $dH_t = 0$ or in state $(0, 0)$ if $dN_t = dH_t = 1$. The expected social loss ΔW_{t+dt} is also that specified in (29). In the absence of jumps, the dynamics of W over $[t, t+dt)$ is given by

$$dW_t + (j\mu - ru_t - \lambda_j \Delta W_{t+dt})\, dt = 0. \tag{30}$$

Under the alternative policy, the dynamics of pool value W' over the same interval is given by

$$dW'_t + (j(\mu + B) - \lambda_j(1+\epsilon)\Delta W_{t+dt})\, dt = 0. \tag{31}$$

Note that $dW_t = \left(1 + v'_j(u_t)\right) \dot{u}_t\, dt \geq 0$ since $v'_j(u_t) \geq -1$ and $\dot{u}_t \geq 0$ under the optimal policy. Since

$$\frac{j\mu - ru_t}{\lambda_j} \geq \frac{j\mu - r\gamma_j}{\lambda_j} \geq \frac{j\mu - r\sum_{k\leq j} b_k}{\lambda_j} = \frac{\mu - rB/\left(\epsilon\bar{\alpha}_j\right)}{\alpha_j} > \frac{\mu + B}{\alpha_j(1+\epsilon)},$$

under Assumption 2, we have

$$0 \leq dW_t = \lambda_j \left(\Delta W_{t+dt} - \frac{j\mu - ru_t}{\lambda_j}\right) dt < \lambda_j(1+\epsilon)\left(\Delta W_{t+dt} - \frac{\mu + B}{\alpha_j(1+\epsilon)}\right) dt = dW'_t.$$

When the optimal policy is resumed at date $t + dt$, we have $W'_{t+dt} = W_{t+dt}$. This implies in turn $W_t > W'_t$, as desired.

A.6 Proposition 6

The most general policy specifies that, upon default,[33] the balance is $j \in \{1, \ldots, I - N_t - 1\}$ with probability $\theta'_j \in [0, 1]$ and 0 otherwise. Let p_j be the penalty associated with downsizing to j. Due to limited liability $b_j \leq u - p_j$. Moreover, incentive compatibility requires that the expected penalty be at least b_{I-N_t}

$$\sum_{j=1}^{I-N_t-1} \theta'_j p_j + \left(1 - \sum_{j=1}^{I-N_t-1} \theta'_j\right) u = u - \sum_{j=1}^{I-N_t-1} \theta'_j (u - p_j) \geq b_{I-N_t}. \tag{32}$$

The corresponding expected efficiency loss is

$$\sum_{j=1} \theta'_j \left(\bar{v}_{I-N_t} - v_j(u - p_j)\right) + \left(1 - \sum_{j=1}^{I-N_t-1} \theta'_j\right) \bar{v}_{I-N_t} = \bar{v}_{I-N_t} - \sum_{j=1}^{I-N_t-1} \theta'_j v_j(u - p_j). \tag{33}$$

By the global concavity of the value functions v_j and given that $\overline{v}_j / b_j$ is increasing, we have

$$\frac{v_j(u - p_j)}{u - p_j} \leq \frac{\overline{v}_j}{b_j} \leq \frac{\overline{v}_{I-N_t-1}}{b_{I-N_t-1}}.$$

Hence the efficiency loss is

$$\begin{aligned}
\bar{v}_{I-N_t} - \sum_j \theta'_j v_j(u - p_j) &= \bar{v}_{I-N_t} - \sum_i \theta'_j (u - p_j) \frac{v_j(u - p_j)}{u - p_j} \\
&\geq \bar{v}_{I-N_t} - \frac{\overline{v}_{I-N_t-1}}{b_{I-N_t-1}} \sum \theta'_j (u - p_j) \\
&\geq \bar{v}_{I-N_t} - \frac{\overline{v}_{I-N_t-1}}{b_{I-N_t-1}} (u - b_{I-N_t}) \\
&= \bar{v}_{I-N_t} - \theta \bar{v}_{I-N_t-1},
\end{aligned}$$

where we have used (32). Now we can attain this lower bound by choosing $\theta'_j = 0$ for all j except $\theta'_{I-N_t-1} = \theta$. This is just the stochastic liquidation policy, as desired.

[33] No liquidation can occur strictly between defaults as it is not optimal to penalize the bank when it is monitoring. Given the bank's impatience, any transfer to investors after the pool balance has reached j can be matched with a lower transfer at the time the balance hits j, with the same effect on monitoring.

A.7 Proposition 7

From the bank's integrated promise-keeping constraint (25-26) along the optimal path, we know that for all $t \leq \tau$

$$u_t = u^* + \int_0^t (ru_s + B(I - N_s)/\epsilon)\, 1_{\{u_s < \gamma_{I-N_s}\}}\, ds - \int_0^t h_s^1\, dN_s - \int_0^t h_s^2\, dH_s,$$

where we have used the fact that $h_t^1 + (1 - \theta_t)\, h_t^2 = b_j$ on $\{I - N_t = j\}$ under a binding constraint. Since $h_t^1 = b_j - [b_j + b_{j-1} - u_t]^+$, this yields

$$u_t = u^* + \int_0^t (ru_s + B(I - N_s)/\epsilon)\, 1_{\{u_s < \gamma_{I-N_s}\}}\, ds - \int_0^t b_{I-N_{s-}}\, dN_s + \xi_t, \tag{34}$$

with ξ_t defined as in Equation (28). Evaluating (34) at $t = \tau$ with $u_\tau = 0$ and $N_\tau = I$, we get

$$0 = u^* + \int_0^\tau (ru_t + B(I - N_t)/\epsilon)\, 1_{\{u_t < \gamma_{I-N_t}\}}\, dt - \sum_{j \geq I - N_{\tau-}} b_j + \xi_\tau, \tag{35}$$

since $I - N_{\tau-}$ is the balance of the pool prior to liquidation.

The sponsor maximizes its profit since by construction

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

The trust's expected cost over the life of the transaction is, using (35),

$$\begin{aligned} u^* + v^* + E\left[\int_0^\tau (ru_t + B(I - N_t)/\epsilon)\, dt + \xi_\tau \right] &= v^* + E\left[\int_0^\tau \delta_t\, dt + \sum_{j \geq I - N_{\tau-}} b_j \right] \\ &= E\left[\int_0^\tau (I - N_t)\, \mu\, dt + \sum_{j \geq I - N_{\tau-}} b_j \right], \end{aligned}$$

which is equal to its expected benefit under the risk retention scheme. Hence the trust breaks even.

B Robustness

In the limiting case $r = 0$, it makes no sense to defer payments following default as the bank is infinitely farsighted. This indicates that the bank is paid continuously. Investors lose an intrument — the stick of stochastic liquidation — and are better off letting the bank hang on to $\gamma_j = \sum_{k \leq j} b_k$, the highest possible target under balance j. There is no risk of private benefit diversion since the bank enjoys the sum of all monitoring rents until extinction of the pool. In the absence of holding costs, the value of the pool is now maximized at all points in time and the first-best is attained. Thus, at $t = 0$, the bank gets an expected payoff of $u^* = \sum_{j \leq I} b_j = (I/\bar{\alpha}_I)(B/\epsilon)$ and the investor lends $v^* = (I/\bar{\alpha}_I)(\mu - B/\epsilon)$. Note that one could still define value functions $v_j(u)$ over $[b_j, \gamma_j]$, but this would be of no practical interest since the stochastic liquidation and probation regions are no longer explored under the optimal plan.

A non-zero liquidation value substantially complicates the analysis but does not change the qualitative properties of the model. Consider first the case when the pool's liquidation value κ_j is small, i.e., lower than the continuation value $\bar{v}_j + b_j$ net of the social value of bank's performance $b_j \left(1 + v_j'(b_j)\right)$:

$$\kappa_j \leq \bar{v}_j - b_j v_j'(b_j). \tag{36}$$

Then, all controls are the same as before. Investors' improved outside option contributes to raising their continuation payoff in the stochastic liquidation region, which by affecting the smooth pasting condition (24) may indirectly lower the bank's target. Investors can now afford to tighten the target and reduce fees accordingly because the presence of recoveries renders the bank's highest level of performance less desirable.

In the second case, when (36) is violated, investors have an even higher outside option and become less tolerant vis-à-vis poor performance. The liquidation threshold shifts to the right from $b_j + b_{j-1}$ to $b_j + \beta_{j-1}$, where $\beta_j > b_j$ is the payoff level implicitly defined by $\kappa_j = v_j(\beta_j) - \beta_j v_j'(\beta_j)$. The difference now is that it is efficient to ask the bank to manage the residual loans starting from a higher performance level, β_{j-1} instead of $u - b_j$, if the pool is maintained. This is without prejudice to the bank, which in this event suffers from a lower utility loss. The stochastic liquidation region shifts to the right from $[b_j, b_j + b_{j-1})$

to $\left[\beta_j, b_j + \beta_{j-1}\right)$,[34] indicating that high liquidation values contribute to shrinking the bank's permissible payoffs further.

We have assumed full commitment. When renegotiation is possible, parties may agree to some Pareto-superior outcome. If v_j has a positive slope, both the investor and the bank would like to restart the contract upon default. This restricts the efficient frontier of the payoff possibility set to states (j, u) such that $v_j'(u) \leq 0$. Renegotiation raises the bank's minimum payoff from b_j to the lowest renegotiation-proof payoff β_j endogenously defined by $v_j'(\beta_j) = 0$, where the value functions v_j are governed by the ordinary differential equations (18)-(19) with linear interpolations taking place on $\left[0, \beta_j\right]$ rather than $[0, b_j]$. When $u < b_j + \beta_{j-1}$, the bank now faces the threat of liquidation because it cannot sustain the full penalty b_j without trying to renogotiate terms. Accordingly, the stochastic liquidation region shifts to $\left[\beta_j, b_j + \beta_{j-1}\right)$. Since the pool is likely to be terminated under higher levels of performance, renegotiation-proofness effectively reduces the profitability of the pool.

[34] Note that in the absence of restrictions about the liquidation values κ_j, the interval could be empty if $\beta_j > b_j + \beta_{j-1}$.

Documents de Travail

360. F. Bec, O. Bouabdallah and L. Ferrara, "The European way out of recessions," January 2012
361. A. Banerjee, V. Bystrov and P. Mizen, "How do anticipated changes to short-term market rates influence banks' retail interest rates? Evidence from the four major euro area economies," February 2012
362. G. Corcos, D. Irac, G. Mion and T. Verdier, "The determinants of intrafirm trade: Evidence from French firms," February 2012
363. C. Glocker, and P. Towbin, "Reserve Requirements for Price and Financial Stability - When Are They Effective?," February 2012
364. C. Altomonte, F. Di Mauro, G. Ottaviano, A. Rungi and V. Vicard, "Global Value Chains during the Great Trade Collapse: A Bullwhip Effect?," February 2012
365. Ph. Andrade and M. Zachariadis, "Global versus local shocks in micro price dynamics," February 2012
366. G. Cette, V. Chouard et G. Verdugo, "Les effets des hausses du SMIC sur le salaire moyen," Février 2012
367. G. Cette and M. de Jong, "Breakeven inflation rates and their puzzling correlation relationships," February 2012
368. S. Dubecq and C. Gourieroux, "Shock on Variable or Shock on Distribution with Application to Stress-Tests," March 2012
369. P. Askenazy, G. Cette and P. Maarek, "Rent building, rent sharing - A panel country-industry empirical analysis," March 2012
370. G. Verdugo, H. Fraisse et G. Horny, " Évolution des Inégalités Salariales en France : le Rôle des Effets de Composition," Mars 2012
371. N. Berardi, Th. A. Eife and E. Gautier, " Optimal Price Setting During a Currency Changeover: Theory and Evidence from French Restaurants," March 2012
372. Ch. A. E. Goodhart, A. K Kashyap, D. P. Tsomocos and A. P. Vardoulakis, "Financial Regulation in General Equilibrium," March 2012
373. M. Bussière, M. Hoerova and B. Klaus, "Commonality in Hedge Fund Returns: Driving Factors and Implications," March 2012
374. C. Glocker and P. Towbin, "The macroeconomic effects of reserve requirements," April 2012
375. E. Gautier and R. Le Saout, "The dynamics of gasoline prices: evidence from daily french micro data," April 2012
376. A. Banerjee and S. Malik, "The changing role of expectations in US monetary policy: A new look using the Livingston Survey," April 2012
377. H. Pagès, "Bank monitoring incentives and optimal ABS," April 2012

Pour accéder à la liste complète des Documents de Travail publiés par la Banque de France veuillez consulter le site : www.banque-france.fr

For a complete list of Working Papers published by the Banque de France, please visit the website: www.banque-france.fr

Pour tous commentaires ou demandes sur les Documents de Travail, contacter la bibliothèque de la Direction Générale des Études et des Relations Internationales à l'adresse suivante :

For any comment or enquiries on the Working Papers, contact the library of the Directorate General Economics and International Relations at the following address :

BANQUE DE FRANCE
49- 1404 Labolog
75049 Paris Cedex 01
tél : 0033 (0)1 42 97 77 24 ou 01 42 92 63 40 ou 48 90 ou 69 81
email : 1404-ut@banque-france.fr