BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2023-3

Last updated: August 9, 2023

Mandatory Retention Rules and Bank Risk

by Yuteng Cheng

Banking and Payments Department
Bank of Canada
ycheng@bankofcanada.ca

DOI: https://doi.org/10.34989/swp-2023-3 | **ISSN** 1701-9397

Acknowledgements

I would like to thank Dean Corbae, Erwan Quintin, Roberto Robatto and Jean-François Houde for invaluable advice and support. I also thank Dmitry Orlov, Rishabh Kirpalani, Paolo Martellini, Timothy Riddiough, Briana Chang, Donald Hausch, Anson Zhou, Jason Choi, Marco Duarte, Yunhan Shin, Jingnan Liu and seminar participants at the University of Wisconsin–Madison for helpful comments. All errors are mine. The views expressed in this paper are those of the author and not necessarily the views of the Bank of Canada.

Abstract

This paper studies, theoretically and empirically, the unintended consequences of mandatory retention rules in securitization. The Dodd-Frank Act and the EU Securitisation Regulation both impose a 5% mandatory retention requirement to motivate screening and monitoring. I first propose a novel model showing that while retention strengthens monitoring, it may also encourage banks to shift risk. I then provide empirical evidence supporting this unintended consequence: in the US data, banks shifted toward riskier portfolios after the implementation of the retention rules embedded in Dodd-Frank. Furthermore, the model offers clear, testable predictions about policy and corresponding consequences. In the US data, stricter retention rules caused banks to monitor and shift risk simultaneously. According to the model prediction, such a simultaneous increase occurs only when the retention level is above optimal, which suggests that the current rate of 5% in the US is too high.

Topics: Financial institutions; Financial system regulation and policies; Credit risk management

JEL codes: G21, G28

Résumé

Cette étude examine, sous un angle théorique et empirique, les conséquences fortuites des règles obligatoires de rétention qui s'appliquent aux titrisations. La loi Dodd-Frank et le règlement de l'Union européenne sur la titrisation imposent tous deux une exigence de rétention du risque de 5 % pour inciter à la sélection et à la surveillance des risques. L'auteur propose d'abord un nouveau modèle montrant que si la rétention renforce la surveillance, elle peut aussi encourager les banques à déplacer le risque. Il fournit ensuite des preuves empiriques de cette conséquence imprévue : les données américaines indiquent que les banques se sont tournées vers des portefeuilles plus risqués après l'entrée en vigueur des règles de rétention enchâssées dans la loi Dodd-Frank. De plus, le modèle permet de faire des prévisions claires et vérifiables sur la politique et ses conséquences. On constate dans les données américaines que les règles de rétention plus strictes ont amené les banques à simultanément surveiller et déplacer le risque. Selon la prévision du modèle, une telle augmentation simultanée ne se produit que lorsque le niveau de rétention est supérieur au niveau optimal, ce qui donne à penser que le taux actuel de 5 % aux États-Unis est trop élevé.

Sujets : Gestion du risque de crédit; Institutions financières; Réglementation et politiques relatives au système financier

Codes JEL : G21, G28

1 Introduction

Securitization, especially of home loans, played an important role in the global financial crisis. Becasue they were packaging and selling downside risk to investors, banks had little incentive to collect and monitor borrower risk (e.g., Gorton and Pennacchi (1995); Parlour and Plantin (2008); Mian and Sufi (2009); Keys et al. (2010)). In response to banks' role in contributing to the crisis, regulators introduced mandatory retention policies. Section 941 of the Dodd-Frank Act and Article 2(1) of the EU Securitization Regulation both imposed a minimum 5% retention rule to better align the incentives of financial intermediaries and asset-backed securitity (ABS) investors. The idea is that by retaining some risk, banks have skin in the game and will therefore invest in screening and monitoring borrowers.[1]

This paper examines, both theoretically and empirically, the impact of retention polices such as those in Dodd-Frank. To accomplish this, I address two key related questions. First, I investigate how mandatory retention affects banks' behaviors in securitization in general. Second, I explore whether the current regulatory level of retention is optimal. My theoretical model demonstrates that retention strengthens monitoring as intended, but beyond a certain threshold, banks start to shift risk, leading to unintended consequences. My empirical analysis of Dodd-Frank's stricter retention rules shows that banks both increased monitoring and shifted risk after the rules were implemented, suggesting that the unintended consequence is not just a theoretical possibility. The findings indicate that the current US retention rate of 5% is too high, as it may exacerbate risk-shifting behavior among banks.

In the model, upon choosing a project to fund, the bank securitizes and sells to investors a fraction of the project, subject to a retention requirement. The bank's credit riskiness is determined by a two-dimensional moral hazard friction. One dimension is *risk shifting* (Jensen and Meckling, 1976), also known as *asset substitution*, which is the opportunity for a bank to replace a high-net present value (NPV) good project with a low-NPV bad project that yields higher private returns if it succeeds (gambling). The bad project thus contains higher credit risk. While the downside risk will be absorbed by debtholders, the private returns go to

[1]Prior to the introduction of mandatory retention rules, securitizers frequently maintained an economic interest in their securitizations, often in the form of first-loss tranches. This was typically driven by several factors, including the mitigation of information asymmetry through the sale of information-insensitive tranches, signaling quality to the market, and regulatory arbitrage, particularly by holding high-rated tranches (Erel, Nadauld, and Stulz (2013)). The extent of retention could vary depending on specific loan characteristics and individual banking institutions (Chen, Liu, and Ryan (2008)). Furthermore, evidence presented by both Flynn, Ghent, and Tchistyi (2020) and Furfine (2020) suggests that the risk retention rules appear to be binding in the context of commercial mortgage-backed securities (CMBS), implying that prior to the crisis, banks commonly retained less than the mandated 5%.

shareholders if the investment strategy pays off. Hence, risk is shifted toward debtholders. The other dimension is *costly monitoring effort* after the securitization stage, which maintains the quality of the project. In practice, monitoring includes collecting payments, renegotiating, and working closely with the trustee representing investors' interests. The bank cannot commit to choosing the good project as well as the monitoring effort. Mandatory retention interacts with the two-dimensional moral hazard. At a low retention ratio, the bank has little incentive to monitor. However, it may still choose to invest in the good project if it is easier to securitize and market rates are favorable. An (exogenous) increase in retention rates causes an increase in monitoring.[2] But given that the bank has limited liability and it is costly to hold a larger share of assets on its balance sheet, the bank is now more likely to invest in the risky project. Consequently, at a high retention ratio, the bank monitors but also increases risk-taking.

With the trade-off, I am able to characterize the socially optimal retention ratio, which I show is strictly positive. Mandatory retention has three effects: it induces monitoring, it may also encourage risk shifting, and it reduces the gain from trade in securitization. The welfare is maximized when good projects are selected and monitoring is incentivized while avoiding risk shifting and excessive regulation of securitization activities. The empirical implication is that if the choice of retention is optimal, we should observe an increase in monitoring but not risk shifting. Conversely, if we observe a simultaneous increase in monitoring and risk shifting, it indicates that the current retention ratio is too high.

The empirical analysis employs the Dodd-Frank Act as a quasi-experiment and uses difference-in-difference estimations to examine the behavior of US bank holding companies (BHC) who are securitizers and also servicers.[3] The final rule of mandatory retention in Dodd-Frank was implemented at different times for different types of securitization. For residential mortgage-backed securities (RMBS), the center of the Dodd-Frank Act, the implementation date was December 2015, while for other securitization categories, it was one year later. I use two risk measures to assess risk shifting and monitoring. The first measure is the ratio of risk-weighted assets

[2] In Appendix B I show that replacing 'monitoring' with ex ante 'screening' does not impact the mechanism. Specifically, increasing screening effort leads to an increase in the upside return of the project, and the expected value of the gambling project increases more than that of the good project. In this context, screening and gambling complement each other.

[3] In my sample, those securitizers keep the servicing rights, meaning that they manage the day-to-day operations of the loan, such as collecting payments from borrowers, handling customer service, and managing escrow accounts. They monitor the performance of the loans, including tracking payments, handling delinquencies, and managing the foreclosure process if necessary. They also oversee loss mitigation processes, such as loan modifications or short sales. While residential mortgage-backed securities (RMBS) tend to concentrate on borrower screening, the role of monitoring – particularly by servicers – is equally crucial. Given this intensive monitoring involved in RMBS, the fundamental mechanism of my model is indeed applicable.

(RWA) to total assets (Furlong, 1988), with an increase in this ratio indicating a risk-increasing change in a bank's asset portfolio. I find that RMBS issuers significantly increased this ratio by 2 percentage points after the effective date of the retention rules on RMBS, suggesting a move toward riskier investment strategies.

The second measure is the delinquency rate, which is the fraction of non-performing outstanding loans. I show that the delinquency rate of RMBS securitizers decreased by an average of 0.3 percentage points after the mandatory rules' implementation, indicating that loans became safer ex post. To account for the longer time it takes for delinquency to occur, an alternative difference-in-difference setup is used to compare RMBS-only securitizers and banks that are mortgage sellers in an extended sample period. Both ordinary least squares (OLS) and propensity score matching yield similar results. The higher ex ante risk of the loans represented by higher wisk-weighted assets ratio and their better ex post performance are consistent with banks exerting more effort to screen and monitor borrowers after the implementation of the retention rules. Overall, there was a simultaneous increase in bank screening/monitoring and risk shifting, and according to the model, this happens only when the current rate of 5% is overshooting.

Lastly, I extend the model to discuss the implications of my results for the optimal retention form. In the final rule of the retention requirements, banks can retain either a horizontal interest, consisting of the most subordinated tranches, a vertical interest in each class of ABS tranches, or a combination of both. I show first that the trade-off between monitoring and gambling exists under both retention forms. Second, the horizontal component generates a higher expected value for the bank when the additional capital requirement brought by subordinated tranches is not a constraint. Finally, I show that at a fixed retention ratio, horizontal retention leads to greater borrower monitoring by the bank after securitization, but also increases the likelihood of over-regulation and risk shifting. This is due to the fact that the larger fraction of the loan retained under horizontal retention effectively resembles an increased retention ratio. From a welfare perspective, the choice of optimal retention form presents yet another trade-off. Ultimately, the findings of this paper have important implications for regulators and practitioners in the banking and financial sectors as they seek to balance these competing objectives in designing policies for securitization.

Related literature. This paper connects several different strands of literature. The first literature studies how securitization negatively affects banks' traditional roles. Theoretically, Pennacchi (1988), Gorton and Pennacchi (1995), Petersen and Rajan (2002), and Parlour and Plantin (2008) argue that securitization leads to a decline of the originating bank's screening and monitoring incentives. Empirically, Keys et al. (2010) and Purnanandam (2011) show that securitization led to lax screening standards of mortgages; Piskorski, Seru, and Vig (2010) and Agarwal et al. (2011) provide evidence that securitization damaged servicing of loans, in particular renegotiation of delinquent loans. I contribute to this literature by analyzing the relationship between securitization and risk shifting.

The second literature discusses the optimal retention form, that is, which tranches banks should retain in response to retention requirements. Fender and Mitchell (2009) and Kiff and Kisser (2014) discuss the optimality of equity and mezzanine tranches in maximizing screening efforts. Pagès (2013) finds that to implement optimal delegated monitoring by the bank, the securitization scheme should use a cash reserve account rather than retention of the residual interest. Malekan and Dionne (2014) study the optimal contract with regard to retention in the presence of moral hazard in lender screening and monitoring. In those papers, the retention requirement is fixed. Instead, I study the optimal requirement, and my results hold under different retention forms. The optimal retention form is also discussed under the broader concept of moral hazard.

This paper is also related to the large literature on risk shifting uncovered by Jensen and Meckling (1976). For an early literature review, see Gorton and Winton (2003). Keeley (1990), Demsetz, Saidenberg, and Strahan (1996), and Repullo (2004) demonstrate the negative first-order effect of profitability on risk shifting. The focus of this literature is mainly capital requirements. This paper instead sheds light on risk shifting and retention requirements in securitization.

Moreover, this paper contributes to the relatively small empirical literature on the impact of retention rules. Furfine (2020) shows that after the implementation of the retention rules, loans in the commercial mortgage-backed securities (CMBS) market become safer, as measured by indexes such as interest rates, loan-to-value ratios, and income-to-debt-service ratios. In a similar manner, Agarwal et al. (2019) find that underwriting standards in the CMBS market are tighter after the implementation of the final rules. The results in the two papers supplement my empirical findings on banks' monitoring and screening behavior, but they do not capture the risk-shifting aspect.[4]

Lastly, there is a literature about signaling private information through retention, pioneered by Leland and Pyle (1977) and DeMarzo and Duffie (1999). Guo and Wu (2014) argue that mandatory retention may deteriorate the adverse selection problem because it prevents issuers from using the level of retention as a signal. On the other hand, Flynn, Ghent, and Tchistyi (2020) show that banks can in fact signal through the retention structure of vertical and hori-

[4]In a related paper, Sarkisyan and Casu (2013) show that retained interests increased bank insolvency risk before the crisis.

zontal interests. The retention policy is fixed in those papers. This paper examines the optimal policy but does not contain signaling. In Chemla and Hennessy (2014), screening is combined with a follow-up possible signaling process through junior tranches, and optimal retention form is also discussed. But in their paper the retention requirement is fixed and there is no risk shifting.

Layout. The paper proceeds as follows. Section 2 describes the institutional background of securitization and risk retention. Section 3 introduces the model, and Section 4 tests its predictions. Section 5 extends the model, and Section 6 concludes the paper.

2 Institutional Background

In a standard securitization process, a loan originator determines whether a borrower qualifies for a loan and, if so, the interest rate of the loan. The originator sells it to an issuer (sometimes referred to as sponsor), who brings together the collateral assets from originators for the asset-backed security.[5] The issuer pools assets together and sells them to an external legal entity, often referred to as a special-purpose vehicle (SPV). The structure is legally insulated from management. The SPV then issues security, dividing up the benefits (and risks) among investors on a pro-rata basis. The issuer usually keeps the servicing rights, that is, the responsibility for managing payments and working closely with the trustee, who represents investors. Credit enhancements are also provided by the issuer to protect investors from potential losses on the securitized assets, the most common forms being subordination, overcollateralization, and excess spread.

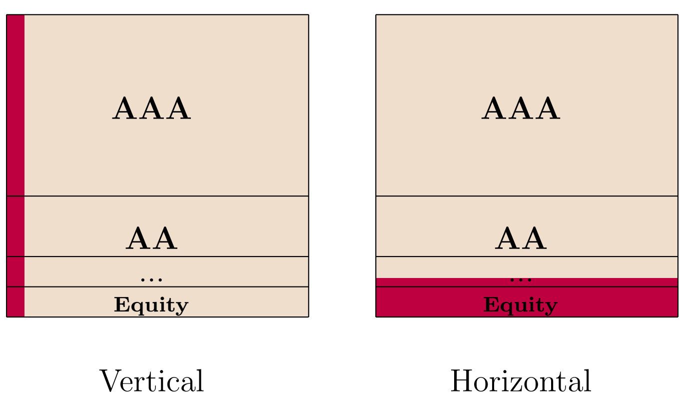

Figure 1: Retention options

[5]The banks I study simultaneously serve as issuers and originators of the portfolio of securitized assets.

In October 2014 after the crisis, to better align the incentives involved in the securitization process, the SEC, FDIC, Federal Reserve, OCC, FHFA, and HUD adopted a final rule (the Final Rule) implementing the requirements of Section 15G of the Exchange Act, which was added pursuant to Section 941 of the Dodd-Frank Act. The Final Rule requires an issuer of ABS to retain at least 5% of the credit risk related to that securitization and restricts the transfer, hedging, or pledge of the risk that the sponsor is required to retain. The issuer must retain either an eligible *vertical interest* (an interest in each class of ABS interests issued as part of the securitization), *horizontal interest* (the issuer holds the most subordinated claim to payments of both principal and interest transactions), as shown in Figure 1, or a *combination of both* so long as the combined retention is not less than 5% of the fair value of the transaction. For the eligible horizontal interest option, the amount of the required risk retention must be calculated under a fair value approach under generally accepted accounting principles (GAAP). The Final Rule came into effect in December 2015 for RMBS and December 2016 for other ABS. A similar retention rule has been introduced by the EU, covered by Article 2(1) of the Securitization Regulation.

Some exemptions exist for particular categories of securitization in the Dodd-Frank Act. For example, ABS or the pooled assets that have the benefit of government guarantees are exempted; sponsors of securitization pools that are solely composed of qualified residential mortgages, as defined by the Consumer Financial Protection Bureau (CFPB) under the Truth in Lending Act, are not required to retain any risk; and collateralized loan obligation (CLO) managers are not subject to risk retention due to a court ruling in 2018. Beyond that, Section 15G permits the agencies to adopt other exemptions from the risk retention requirements for certain types of ABS transactions.

3 The Model

This section presents the main model. I illustrate how mandatory retention in the securitization process affects banks' decisions and market outcomes. I use uppercase letters to represent banks' decision variables, while parameters are denoted using lowercase and Greek letters. The model of this section assumes that banks use a vertical retention form for exposition and simplicity (see Section 2 for a discussion of retention form), but Section 5 shows that the main results are unchanged when banks can use both forms of retention.

3.1 Model Setup

Model overview. Consider an economy with a bank and multiple investors, both assumed to be risk-neutral. The economy lasts for four dates: $t \in \{0, 1, 2, 3\}$. At $t = 0$, the bank needs to invest one unit of money in a project using equity e and debt $d = 1 - e$. The bank chooses between two projects: a good project and a bad (gambling) project, the details of which will be defined later. At time 1, the bank securitizes a fraction of its project to investors. The securitization process is regulated by mandatory retention requirements, which will also be specified below. At time 2, the bank chooses whether to monitor its project $M \in \{0, 1\}$, where $M = 1$ denotes monitoring and $M = 0$ denotes no monitoring. The monitoring effort affects the credit risk of the project as a whole, not just the portion the bank retains. At time 3, returns are realized. The timeline of the model is in Figure 2.

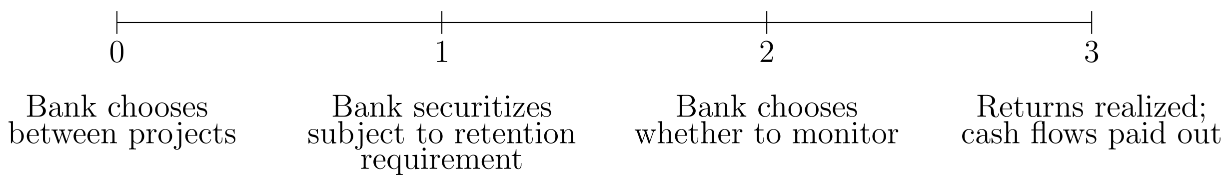

Figure 2: Timeline of the events

Risk-shifting problem. At time 0, the bank faces a risk-shifting problem (Jensen and Meckling, 1976) when it chooses which project to invest in: $K \in \{g, b\}$. The term "project" can be interpreted as the bank's portfolio of investments. The expected return of a project depends on monitoring, but for any level of monitoring, the good project g has higher expected returns. Meanwhile, the bad project b yields higher private returns if the gamble pays off. More specifically, given a monitoring level $M \in \{0, 1\}$, the project payoffs are

$$\underset{\text{good project}}{\text{Payoff}} = \begin{cases} R_g & \text{w/prob } q_g(M) \\ r & \text{otherwise} \end{cases}, \qquad \underset{\text{bad project}}{\text{Payoff}} = \begin{cases} R_b & \text{w/prob } q_b(M) \\ r & \text{otherwise} \end{cases}$$

where $R_b > R_g > 1 > r$, and

$$q_g(M) = p_g - (1 - M)\Delta_g,$$
$$q_b(M) = p_b - (1 - M)\Delta_b,$$

with $p_g > p_b$. To clarify, when the bank opts to monitor ($M = 1$), the success probabilities for the projects are just p_g and p_b, and the project payoffs are

$$\underset{\text{good project}}{\text{Payoff } (M{=}1)} = \begin{cases} R_g & \text{w/prob } p_g \\ r & \text{otherwise} \end{cases}, \qquad \underset{\text{bad project}}{\text{Payoff } (M{=}1)} = \begin{cases} R_b & \text{w/prob } p_b \\ r & \text{otherwise} \end{cases}.$$

When the bank chooses not to monitor ($M = 0$) at time 2, the probability of success for a project drops by Δ_K, and the project payoffs are

$$\underset{\text{good project}}{\text{Payoff }(M{=}0)} = \begin{cases} R_g & \text{w/prob } p_g - \Delta_g \\ r & \text{otherwise} \end{cases}, \qquad \underset{\text{bad project}}{\text{Payoff }(M{=}0)} = \begin{cases} R_b & \text{w/prob } p_b - \Delta_b \\ r & \text{otherwise} \end{cases}.$$

I assume $\Delta_g < \Delta_b$, meaning that without monitoring, the good project's success probability decreases less than the bad project's. This implies the bad project's success probability is always lower, $q_b(M) < q_g(M)$ for any M.

I make two further assumptions. First, I assume that if the cash flow r is realized, the bank fails and r can be considered as the liquidation value of the bank. For simplicity, I assume $r = 0$ in this section and relax that assumption in Section 5. And without loss of generality, I normalize $p_g = 1$.[6] The stochastic returns of the two projects are then assumed to satisfy

$$R_b > R_g > p_b R_b > 1.$$

The first inequality corresponds to the fact that the gambling project yields a higher payoff in the success state, and the second inequality corresponds to the fact that the good project has a higher expected return at any monitoring level.[7] The last inequality states that both projects have positive NPV.

If the bank is solvent at time 2, it repays d to debtholders. The bank has limited liability. When the gamble fails, the cost is borne by debtholders. I make the following assumption that the amount of debt the bank needs to repay, d, is not too small.

Assumption 1 $d > \frac{R_g - p_b R_b}{1 - p_b}$.

This assumption guarantees that risk shifting is possible, at least when the bank retains everything on the balance sheet, that is, $p_b(R_b - d) > R_g - d$.

Securitization and mandatory retention. At time 1, the bank securitizes an exogenous fraction α of its project, which is then repackaged into multiple securities or "tranches" with varying seniorities. The most junior tranche will be the first to absorb losses, followed by subsequent tranches in order of seniority.[8] Throughout this paper, the securitized fraction of the project will be referred to as the "pool."

[6] This assumption implies that when $m = 1$, the good project is riskless.

[7] To see this, note that $R_g > p_b R_b$ implies $(1 - \Delta_g)R_g > (p_b - \Delta_b)R_b$.

[8] It should be noted that the design of these securities, or subordination in practice, is considered exogenous and will not be the focus of this paper. See Mitchell (2004) for an overview of tranching and related literature on security design.

The securitization process is regulated by *mandatory retention requirements*: the bank chooses to retain Z fraction of the tranches such that the retained tranche value is *no less than* θ fraction of the market value of the pool. The policymaker sets θ. In the vertical retention we focus on in this section, the bank simply holds each class of the tranches; hence, the retention constraint is

$$Z \geq \theta,$$

that is, the bank needs to retain at least θ fraction of the pool.

Competitive investors observe the bank's project choice K and retention choice Z, and bid a price (schedule) P.[9] The total monetary benefit of securitization for the bank at time 1 is λP, where $\lambda > 1$ is the gain from trade. It can be interpreted, for instance, as lower funding costs in extending new loans, or new investment opportunities, which I do not model preciesely, or an increase in reported profit that is related to the compensation of the CEO. Moreover, $\lambda > 1$ is equivalent to the bank discounting future cash flows at a higher rate than investors (e.g., DeMarzo and Duffie, 1999). One underlying assumption here is that the cash flows generated from securitization are primarily used either for future business operations or as dividends to shareholders. This assumption can be relaxed to allow for a partial allocation towards the repayment of existing debts (d), but not for the complete repayment.[10]

Monitoring. At time 2, the bank chooses whether to monitor: $M \in \{0, 1\}$. The influence of this choice on stochastic returns is presented above. The monitoring process incurs a cost, denoted as c, applicable to both project types.

Equilibrium definition. I present the definition of equilibrium in this model.

Definition 3.1 (Equilibrium.) *The equilibrium consists of the bank's project choice K at time 0; the bank's retention choice Z at time 1; price P of the tranches sold to investors; the bank's monitoring choice M at time 2, such that P reflects seniorities and project and retention choices. Investors break even and the bank maximizes its value sequentially subject to retention constraint at time 1.*

[9]The project choice being observable to securitization purchasers is in accordance with ABS-offering prospectuses that provide detailed information on the underlying collateral assets, as governed by Regulation AB from 2005 and reinforced in Dodd-Frank. And the transparency of the retention choice is governed by the retention rules.

[10]If all cash flows were used solely for debt repayment, the bank would monitor more when there is less retention, due to $\lambda > 1$. The intended consequences of mandatory retention rules cannot be achieved. Therefore, I exclude this possibility from my analysis.

In the subsequent sections, for the sake of clarity and ease of exposition, the subscript K in p_K, R_K and Δ_K will be omitted, provided that does not result in any ambiguity.

3.2 Monitoring decision

The model is solved backwards. In this section, I solve for the bank's monitoring decision at time 2, after the securitization stage. Given the project and retention choices (K, Z), the optimal monitoring decision maximizes the bank's residual profits net of monitoring costs:

$$\Pi_2(K, Z) = \max_{M \in \{0,1\}} \; q_K(M)\Big((1-\alpha)R + Z\alpha R - d\Big) - cM.$$

In the expression, when the investment is successful, $(1-\alpha)R$ is the value of the non-securitized part of the project, $Z\alpha R$ is the value of the tranches retained on the bank's books, and d is the amount of debt the bank repays being solvent. I present the solution.

Proposition 3.1 *Given (K, Z), the bank chooses $M = 1$ if and only if*

$$Z \geq \bar{Z}_K \equiv \frac{\frac{c}{\Delta} + d - (1-\alpha)R}{\alpha R}.$$

Proof: All the proofs are presented in Appendix A.

Here I assume that the bank monitors when there is a tie. The bank monitors if the loan retained on the balance sheet is large enough, that is, if there is "skin in the game."[11] The threshold $\bar{Z}_K$ is increasing in the inverse of the efficiency of the monitoring technology, measured by $\frac{c}{\Delta}$, and total debts d it owes: when the monitoring technology is highly efficient, the bank will advance its monitoring decision to increase profits. Conversely, if the bank owes significant debts (higher leverage), it may hesitate to monitor because the benefits of being solvent may be outweighed by the costs of paying off the debts. The threshold is decreasing in the project return R_K. If the bank chooses the bad project (with $R_b > R_g$ and $\Delta_b > \Delta_g$), it has more incentive to monitor. Put another way, $\bar{Z}_g > \bar{Z}_b$. Monitoring and gambling are hence complements.

3.3 Securitization problem at time 1

In vertical retention, investors purchase from the bank $1 - Z$ fraction of the pool and share the default risk of the underlying project evenly in terms of seniority with the bank. If the bank

[11]To further illusatrate the point in footnote 10, when cash flows generated from securitization (which will be $\lambda\alpha ZR$ as shown later) are used entirely for debt repayment, one can show that contrary to the result above, the bank monitors if and only if Z is below some threshold.

goes bankrupt, the debtholders will take over the tranches. Although the transaction takes place before the monitoring stage and the bank cannot commit to monitoring, investors who observe Z can later infer the bank monitoring decision by comparing it to $\bar{Z}_K$ and pay the corresponding adjusted price. Specifically, for a given project, the schedule of prices investors offer based on Z and rational expectations is

$$P(Z|K) = \begin{cases} (1-Z)\alpha pR & \text{if } Z \geq \bar{Z}_K, \\ (1-Z)\alpha(p-\Delta)R & \text{if } Z < \bar{Z}_K. \end{cases}$$

The expected value of the pool will be αpR if there is subsequent monitoring and $\alpha(p-\Delta)R$ if there is not.

The bank's Bellman equation at time 1 is expressed as follows:

$$\Pi_1(\theta|K) = \max_Z \ \lambda P(Z|K) + \Pi_2(Z|K)$$
$$\text{s.t. } Z \geq \theta.$$

Here, the total monetary benefits of securitization are represented by the product of λ and $P(Z|K)$. The retention requirement is captured by the inequality constraint. In the vertical retention, the bank only needs to retain a fraction larger than θ of the pool. I first characterize the shape of the bank's objective function as a function of Z.

Proposition 3.2 *The bank's objective function $\lambda P(Z|K) + \Pi_2(Z|K)$ is upper semicontinuous at $\bar{Z}$. Moreover, it is linear and decreasing in Z on $[0, \bar{Z}_K)$ and $[\bar{Z}_K, 1]$.*

To make monitoring indeed a friction, I assume for each project K the following holds.

Assumption 2 *$\lambda\big(\Delta R - pd\big) < c$ for both projects.*

To have enough incentive to monitor, the bank must retain at least $\bar{Z}_K$ of the project. While monitoring improves project quality, it comes at a cost: lowered securitization revenue due to retention and the cost of monitoring. This assumption guarantees that in equilibrium, the bank chooses zero retention in the absence of mandatory retention rules. However, if this assumption is violated, the bank always chooses to retain $Z = \bar{Z}_K$ and monitor. In this case, there is no need for regulation to address the misaligned incentive of monitoring. The left side of Figure 3 illustrates the bank's objective function under this assumption, with the second piece of the function being steeper than the first.

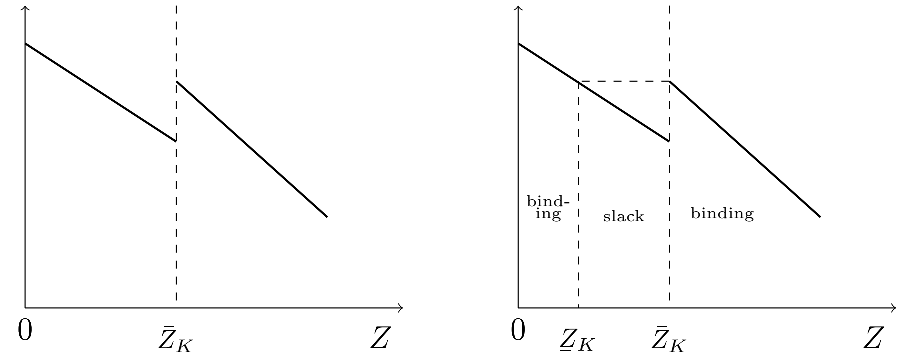

Figure 3: The bank's objective function and retention constraint

We are now ready to impose the retention constraint. Let $\underline{Z}_K$ be such that

$$\lambda P(\underline{Z}_K|K) + \Pi_2(\bar{Z}_K|K) = \lambda P(\bar{Z}_K|K) + \Pi_2(\bar{Z}_K|K).$$

According to the shape of the objective function, this $\underline{Z}_K$ exists and is smaller than $\bar{Z}_K$, as shown in the right part of Figure 3. In fact, the formula of $\underline{Z}_K$ for project K is

$$\underline{Z}_K = 1 - \frac{(\lambda p - (p-\Delta))(R-d) - (\lambda-1)\frac{c}{\Delta} - c}{(\lambda-1)\alpha(p-\Delta)R}. \tag{1}$$

If the required threshold θ is below $\underline{Z}_K$, the bank chooses the lowest possible retention, implying that the retention constraint is binding. When the threshold θ is between $\underline{Z}_K$ and $\bar{Z}_K$, the objective function is maximized at $\bar{Z}_K$, indicating that the retention constraint is not binding. However, if θ is above $\bar{Z}_K$, the constraint is binding again since the objective function is decreasing. These are summarized by the following result.

Proposition 3.3 (Retention choice.) *The bank's optimal retention choice of Z is*

$$Z(\theta|K) = \begin{cases} \theta & \text{if } \theta < \underline{Z}_K, \\ \bar{Z}_K & \text{if } \underline{Z}_K \leq \theta < \bar{Z}_K, \\ \theta & \text{if } \bar{Z}_K \leq \theta. \end{cases}$$

If there is no retention requirement, the bank will not make any monitoring efforts. The policymaker can set the lower bound θ of the retention requirement to $\underline{Z}_K$ to encourage monitoring from the bank, which we show later to be part of the socially optimal allocation. Since $\underline{Z}_K$ is smaller than $\bar{Z}_K$, the policymaker only needs a small lower bound of θ.

Proposition 3.4 (More retention leads to more monitoring.) *Given a project K, the policymaker can set $\theta \geq \underline{Z}_K$ to implement monitoring. In particular, if the policymaker sets*

$$\theta \geq \underline{Z}_g \equiv 1 - \frac{(\lambda - (1-\Delta_g))(R_g - d) - (\lambda-1)\frac{c}{\Delta_g} - c}{(\lambda-1)\alpha(1-\Delta_g)R_g},$$

the bank always monitors.

For a given project K, under the optimal retention choice, the bank's expected value $\Pi_1(\theta|K)$ is graphed in Figure 4. It is constant on $[\underline{Z}_K, \bar{Z}_K]$ because of the slack retention constraint. The second part of the result comes from the fact that the good project has a larger $\underline{Z}_K$.

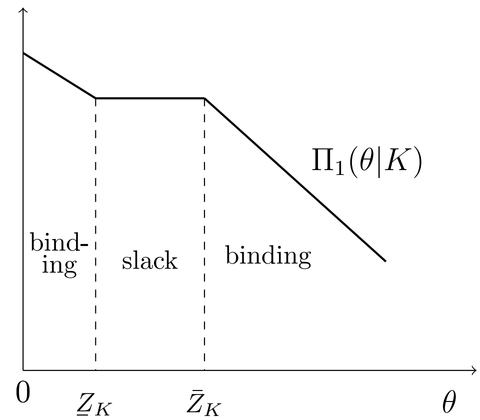

Figure 4: The bank's expected value under the optimal retention choice

3.4 Risk shifting at time 0 and equilibrium

In this section, I complete the equilibrium characterization by studying the bank's risk-shifting motives in the presence of limited liability, which imposes more costs on creditors in bad outcomes. I show that mandatory retention increases the bank's risk-shifting propensity, defined by the difference between the values of gambling and being good concerning project choice.

At time 0, for a required retention ratio θ, let the bank's expected value of choosing the good project be $V_g(\theta) \equiv \lambda P(Z_g(\theta)|g) + \Pi_2(Z_g(\theta)|g)$. The retention $Z_g(\theta)$ comes from Proposition 3.3. Similarly, if the bank shifts risk, its value changes to $V_b(\theta) \equiv \lambda P(Z_b(\theta)|b) + \Pi_2(Z_b(\theta)|b)$. The bank hence solves

$$\max_K V_K(\theta).$$

To solve this, I define the gain from risk shifting as

$$G(\theta) \equiv V_b(\theta) - V_g(\theta),$$

and the bank chooses to gamble if $G(\theta) > 0$. The equilibrium depends on the behavior of this gain function under the different requirements of θ. Note that this function has four kink points since both $V_g(\theta)$ and $V_b(\theta)$ have two: $\underline{Z}_g, \bar{Z}_g$ and $\underline{Z}_b, \bar{Z}_b$, as shown in Figure 4. I make the following assumption so that in equilibrium it is possible for the bank to optimally choose the good project and monitor it at the same time.

Assumption 3 $(\lambda - 1)c\left(\frac{1}{\Delta_g} - \frac{p_b}{\Delta_b}\right) > \lambda[p_b(R_b - d) - (R_g - d)]$.

It guarantees that $V_g(\bar{Z}_g) > V_b(\bar{Z}_b)$: that is, when both retention constraints are slack, if the bank starts to monitor the good project, it obtains a higher value from that project than from gambling. The main result of the paper is the following.

Proposition 3.5 (Equilibrium) *There exist two thresholds $\underline{\theta}, \bar{\theta}$ with $\underline{\theta} < \bar{\theta}$ such that the bank chooses the good project but does not monitor for $\theta < \underline{\theta}$; the bank chooses the good project and makes an effort to monitor for $\theta \in [\underline{\theta}, \bar{\theta}]$; the bank gambles and monitors subsequently for $\theta > \bar{\theta}$.*

The above equilibrium characterization implies that at a high retention ratio θ, the bank starts to gamble. Because the gambling project has lower NPV than the good project, the total output is lower. This is the unintended consequence of mandatory retention rules.

Corollary 3.1 (Unintended consequence of mandatory retention.) *The bank starts to gamble when $\theta > \bar{\theta}$.*

The two possibilities of $V_g(\theta)$, $V_b(\theta)$, and the corresponding $G(\theta)$ are portrayed in Figure 5. At a low retention ratio, the bank chooses the good project because it has favorable market value and hence is easy to securitize; and this is the driving force of $G(\theta)$ being monotonic increasing on some region. As the retention ratio increases, the value of choosing the good project is hurt more because of the higher gain from trade generated. When the retention ratio is high enough, limited liability and higher upside return R_b make the gambling project more profitable.

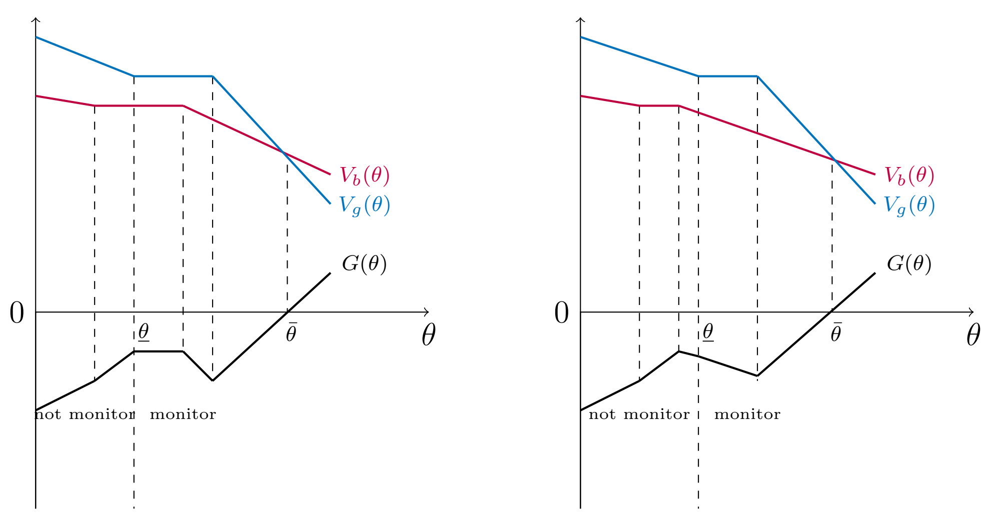

Figure 5: Risk-shifting propensity

It is worthwhile to compare the outcomes of the two other scenarios: selecting the good project without monitoring, and selecting the gambling project with monitoring. I assume that

$$p_b > 1 - \Delta_g.$$

That is, the default rate is lower when there is monitoring despite some risk being shifted. First, this is consistent with the narrative that lax monitoring in securitization leads to bad loans and the global financial crisis. Moreover, it also matches the empirical evidence explored later showing that the mandatory retention rule has caused banks to monitor and gamble simultaneously, but loan delinquencies have dropped.

3.5 Welfare and optimal retention requirement

This section provides the welfare analysis and discusses the optimal retention ratio. The total welfare is defined as the sum of the payoffs of all players[12] as a function of θ. Since investors get zero in expectation, the sum is just the bank's value V^{bank} plus debtholders' expected return $p(\theta)d$, where $p(\theta)$ is the probability of the equilibrium project being successful. In other words,

$$W(\theta) = V^{\text{bank}}(\theta) + p(\theta)d,$$

where

$$V^{\text{bank}}(\theta) = \begin{cases} V_g(\theta) & \text{if } \theta \leq \bar{\theta}, \\ V_b(\theta) & \text{if } \theta > \bar{\theta}, \end{cases}$$

according to Proposition 3.5 and

$$p(\theta) = \begin{cases} 1 - \Delta_g & \text{if } \theta \leq \underline{Z}_g, \\ 1 & \text{if } \theta \in [\underline{Z}_g, \bar{\theta}], \\ p_b & \text{if } \theta > \bar{\theta}. \end{cases}$$

Note that $V^{\text{bank}}(\theta)$ is continuous but $p(\theta)$ is not. After rearrangements, the welfare can be written as

$$W(\theta) = \big(1 + (\lambda - 1)\alpha(1 - Z(\theta))\big)\bar{R}(\theta) - c(\theta),$$

which is the sum of total output and the net gain from securitization less the potential monitoring cost. The shape is shown in Figure 6.

There are jumps at $\underline{Z}_g$ and $\bar{\theta}$ because $p(\theta)$ has jumps at these points. Since securitization brings gains of trade, a higher retention ratio reduces securitization levels and, hence, brings down the total welfare. This is why the welfare is decreasing on each piece, except for $[\underline{Z}_g, \bar{Z}_g]$, on which the retention constraint is slack. In fact, if we impose the following assumption, the welfare is maximized at $[\underline{Z}_g, \bar{Z}_g]$ and maximized when the good project is chosen and monitoring effort is delivered:

[12] That is, equal Pareto weights are attached to each player.

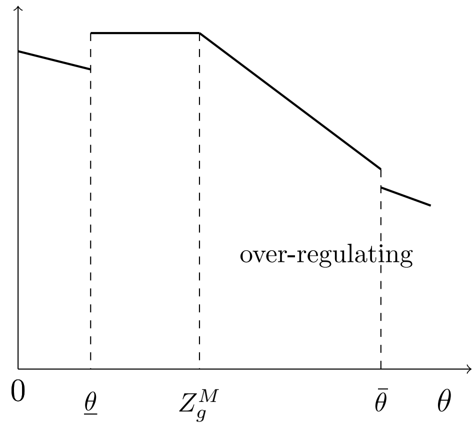

Figure 6: Welfare

Assumption 4 $\Big\{(\lambda-1)[1-\alpha(1-\Delta_g)]+\Delta_g\Big\}R_g-(\lambda-1)d > c+(\lambda-1)\frac{c}{\Delta_g}$.

This assumption is only about the good project; hence, it is easy to satisfy when, say, R_g is large enough. Similar to the benchmark case, the social monitoring cost and loss of gains from trade due to retention are low compared to the additional social benefits generated from monitoring the good project. As λ converges to 1, the above assumption converges to $\Delta_g R_g > c$. Since $\underline{\theta} = \underline{Z}_g$, we have the following result.

Proposition 3.6 (Optimal retention.) *The optimal retention ratio is $[\underline{\theta}, \bar{Z}_g]$. On this range of θ, the bank invests in the good project and monitors, and the retention constraint is slack.*

The optimal retention ratio is interior, meaning that a certain degree of retention is the correct policy. Mandatory retention has three effects: it induces monitoring, it encourages risk shifting if the degree of retention is very high, and it restricts securitization activities and hence reduces the gains from trade. What happens when θ is between $\bar{Z}_g$ and $\bar{\theta}$? In this region, the bank still chooses the good project and delivers monitoring effort according to Proposition 3.5; however, its securitization activities are over-regulated. In summary, welfare is maximized when the good project is selected and just monitored, without securitization activities being overly regulated.

Two remarks are worth making. First, if the policymaker places a 100% Pareto weight on debtholders, total welfare is simply $p(\theta)$, the probability of success. The optimal retention ratio is $[\underline{\theta}, \bar{\theta}]$. Second, the policymaker is imposing an inequality constraint. If this inequality constraint is replaced with an equality constraint, that is, the bank has to retain θ fraction of the pool, the optimal retention is unique: $\theta^* = \bar{Z}_g$. This is the lowest level of retention to incentivize the bank to monitor. And the policymaker does not want to go beyond that to over-regulate securitization. Since $\bar{Z}_g$ is larger than $\underline{\theta}$, $\bar{Z}_g$ can be close to $\bar{\theta}$, above which the bank shifts risk. Therefore, choosing an inequality retention constraint provides the policymaker with more flexibility to balance monitoring and risk shifting.

3.6 Model predictions

This section connects the above results with empirical tests in the next part of the paper. The model provides sharp predictions about the link between policy and unintended consequences, which is summarized in Table 1. In particular, if we observe no changes in monitoring and risk shifting, it implies that the current retention ratio is below the optimal level. Similarly, if we observe a simultaneous increase in monitoring and risk shifting, the current retention ratio is too high. [13]

	Monitor	Shift risk
$\theta < \underline{\theta}$	N	N
$\theta \in [\underline{\theta}, \bar{\theta}]$	Y	N
$\theta > \bar{\theta}$	Y	Y

Table 1: Model predictions

As we will demonstrate in the next section, banks do monitor more and shift more risk after the retention rule's implementation, implying that the current rate of 5% is overshooting.[14]

4 Empirical Evidence

This section tests the model predictions. Since risk and capital management are typically carried out at the highest level, I use US bank holding company (BHC) data from Y-9C forms. A BHC is defined as a securitizer if it reports at least one non-zero outstanding securitization activity in the sample period and has at least two years of existence. From 2001Q2, BHCs were required to detail their securitization activity in their regulatory reports (Schedule HC-S of the Y-9C report). Securitization activity is the outstanding principal balance of assets sold and

[13]Admittedly, since the optimal retention ratio (when the planner places equal weights) $[\underline{\theta}, Z_g^M]$ is embedded in $[\underline{\theta}, \bar{\theta}]$, we cannot distinguish between an optimal retention requirement and an over-regulating retention requirement when monitoring is observed and risk shifting is not in this case.

[14]Note that if Assumption 4 fails, the socially optimal retention ratio is 0. This does not change this implication.

securitized with servicing retained or with recourse or other seller-provided credit enhancements in millions of US dollars. If another bank acquired a bank, I remove the non-survivor from the sample.

I use two measures to document bank risk. The first is the risk-weighted assets (RWA) ratio, which is calculated by dividing the RWA by the total assets of a bank. To determine a bank's RWA, its assets are categorized by risk level and the likelihood of causing a financial loss. In simplified terms, the RWA ratio is computed as follows:

$$\text{RWA ratio} = \frac{\sum_{k=1}^{n} \omega_k A_k}{\sum_{k=1}^{n} A_k},$$

where ω_k denotes the risk weight of asset k and A_k denotes the value of asset k. If the ω_k values are fixed, investing in assets with higher risk weights will increase the numerator while keeping the denominator constant. Therefore, an increase in the RWA ratio indicates that the bank is allocating more resources to riskier projects, or engaging in risk shifting.[15]

The second measure is the securitization delinquency ratio, defined as

$$\text{DEL} = \frac{\text{loans 90 or more days past due} \ + \ \text{(loans charged off)}}{\text{lagged total loans outstanding}}.$$

The delinquency rate measures the fraction of loans outstanding that are non-performing loans; hence, it is an ex post index of bank risk.

I use two difference-in-difference (DID) setups to show a notable rise in the RWA ratio for RMBS securitizers and a decline in the loan delinquency rate following the implementation of the retention rules. The first DID estimation compares US BHC RMBS securitizers to other BHC securitizers from 2015Q1 to 2016Q4, where the mandatory retention rule took effect for RMBS in December 2015, and one year later for all other categories of securitization.[16] However, since delinquency takes place at least 90 days after a loan is generated, a short sample period may not yield accurate estimates. Therefore, I also consider an alternative DID design with a longer horizon (2014–2017) and compare RMBS-only BHC securitizers to non-securitizers who are mortgage sellers.

[15] In the US Basel framework Final Rule, there are two implementation approaches to calculate RWA. The standard approach, which applies to all banks, attaches fixed risk weights to multiple exposure types. In the advanced approach, which applies to banks with consolidated assets greater than \$250 billion, exposures are broadly classified into four categories: retail, wholesale, securitization, and equity. In addition, the internal ratings-based (IRB) formula applies to retail and wholesale exposures. The standard and advanced approaches took effect in January 2015 and January 2014, respectively. Banks with total assets greater than \$250 billion have to calculate RWA using both standardized and advanced approaches and may act differently from the rest of the banks.

[16] While the Dodd-Frank Act was passed in July 2010, the final rule of mandatory retention was, in fact, not agreed upon until October 2014. There was then a two-year delay between the agreement and implementation.

4.1 DID: RMBS securitizers versus other securitizers

In the first DID setup, the focus is the effect of the retention rules on banks' risk-shifting behavior, measured by an increase in the RWA ratio. I use US BHC data from 2015Q1 to 2016Q4 from Y-9C forms.

As mentioned in footnote 15, banks with consolidated assets greater than $250 billion report RWA using two different approaches. I first remove those banks and add them back as a robustness check. There are 52 BHC securitizers and 359 bank-quarter observations. Banks that issue RMBS may also be involved in other categories of securitization businesses. I consider the standard binary treatment as well as a continuous treatment.

In the binary treatment, banks are split into control and treatment groups. The control group contains 12 BHC securitizers that do not issue RMBS. The treatment groups are banks that issue RMBS, but with different issuance shares, defined as

$$Share = \frac{\text{RMBS activities}}{\text{Total securitization activities}}.$$

The RMBS activities, as well as total securitization activities, do not contain loans sold to other institutions or entities.[17] I consider three definitions of the treatment group: (1) $share > 0.5$ every quarter; (2) $share > 0.8$ every quarter; (3) $share = 1$ every quarter. There are 40 banks in total that are RMBS issuers: 37 of them are in group 1, 36 in group 2, and 33 in group 3. Securitization activities and shares are stable across time for RMBS securitizers. There is one single bank that did not issue RMBS until 2016Q3, and its share is below 0.15.I hence include this bank in the control group. After defining the variable D_{it} to equal 1 if bank i is an RMBS securitizer after December 2015 and 0 otherwise, the benchmark model can be expressed as the following fixed-effect panel regression:

$$RWA_{it} = ZD_{it} + X_{it}'\delta + \alpha_i + \gamma_t + \epsilon_{it}, \tag{2}$$

where subscripts it uniquely identify individual observations for bank i in quarter t. The dependent variable is the RWA ratio. I include bank fixed effects to absorb unobservable differences in bank business models and time fixed effects to absorb macro-economic shocks such as quantitative easing. The independent variable γ_t is time fixed effects, X_{it} is the set of bank-specific control variables, α_i is the individual fixed effects, and ϵ_{it} is the error term.

[17] Examples of other institutions and entities are the Federal National Mortgage Association (Fannie Mae) or the Federal Home Loan Mortgage Corporation (Freddie Mac). These government-sponsored agencies, in turn, securitize these loans and are not subject to the mandatory retention. These items are reported separately in Items 11 and 12 in Schedule HC-S.

Standard errors are clustered at the bank level. The coefficient of interest is Z. I estimate (2) with and without bank-specific controls.

To shed light on the dynamics of the average treatment effect before and after the passage of the mandatory retention rule, I replace the dummy D_{it} in specification (2) with leads and lags, which I name as $Before$ and $After$. The new regression model is

$$RWA_{it} = \alpha_i + \sum_{N=2}^{4} Z_{5-N} B_{N,it} + \sum_{M=0}^{3} \alpha_{3-M} A_{M,it} + X_{it}'\delta + C_t + \epsilon_{it}, \tag{3}$$

where B_N denotes N periods before the treatment and A_M denotes M periods after treatment. The quarter before treatment is used as the benchmark. The coefficients Z to α estimate the average changes in banks' risk in the quarters preceding and following the Final Rule's implementation. The aim here is to test whether the effect is isolated to periods occurring only after the onset of the implementation.

The set of bank-specific control variables includes on-balance-sheet and off-balance-sheet items. On-balance-sheet items include some standard explanatory variables in the banking literature, like size (measured by the log of total assets), profitability (measured by return on assets), capital buffer (measured by the difference between banks' regulatory risk-based capital and the minimum required capital ratio), and liquidity ratio (over total assets). Banks of different sizes are regulated with different stringencies. For example, banks with assets of more than \$50 billion are subject to stress tests, must submit resolution plans, and have tighter liquidity requirements. The capital buffer and liquidity buffer are a sign of bank solvency and stability. The off-balance-sheet item I use is securitization activities scaled by total assets.

Table 2: **Summary statistics**

	Control group		RMBS share≥0.8		RMBS share=1	
Variables	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Total assets ($ billions)	66.897	63.088	39.754	58.132	26.591	37.946
RWA ratio	0.780	0.119	0.749	0.092	0.748	0.089
RWA ratio (exclude securitization exposure)	0.776	0.119	0.748	0.091	0.748	0.089
Loan ratio (of total assets)	0.646	0.165	0.691	0.100	0.698	0.089
Home mortgage (of total loans)	0.226	0.151	0.322	0.171	0.325	0.180
Consumer loans (of total loans)	0.168	0.162	0.081	0.099	0.080	0.104
C&I loans (of total loans)	0.232	0.124	0.180	0.102	0.177	0.105
Capital buffer	0.101	0.134	0.138	0.915	0.149	0.982
Deposit ratio	0.419	0.186	0.587	0.112	0.597	0.101
Liquidity ratio	0.207	0.104	0.216	0.075	0.213	0.077
ROA	0.009	0.013	0.006	0.005	0.006	0.005
Securitization-assets ratio	0.142	0.340	0.129	0.161	0.122	0.161
Number of banks	12		40		33	

Notes: The table presents descriptive statistics of the control group and treatment groups from 2015Q1 to 2016Q4. I report treatment groups with an RMBS activity share greater than 0.8 as well as equal to 1. The table contains means and standard deviations of bank characteristics. The RWA ratio is the risk-weighted assets to total assets ratio. ROA is the ratio of net income to total assets. The securitization-assets ratio is the ratio of securitization activities to total assets. Credit enhancements is the ratio of total credit enhancements provided by the securitizer to total assets.

The summary statistics for the treatment and control group are reported in Table 2. I report treatment groups with an RMBS activity share greater than 0.8 as well as equal to 1. The average total assets of banks in the control group is $60.8 billion, which is greater than banks in the treatment group, whose average total assets is under $40 billion. Moreover, as I increase the share of RMBS activities, the average size drops. The average size becomes $26.6 billion when banks only issue RMBS. Banks in both groups participate in securitization activities each quarter at levels about 13% to 14% of their size. Regarding on-balance-sheet items, banks in the treatment group have a slightly higher loan ratio (69% versus 65%) and a higher home mortgage ratio over total loans (32% versus 23%), which is no surprise. Banks in the control group focus more on consumer loans and C&I loans. Banks in both groups hold 21% of their total assets in the form of liquid assets. The average RWA ratio in the full sample is 0.76. Banks in the control group have a higher RWA ratio. The difference between the two groups is around 0.03 (0.78 versus 0.75). I also report the RWA ratio excluding securitization exposure. The numbers do not change much. Though operating at a higher asset level, banks in the control group have more profitable projects than banks in the treatment group (0.009 versus 0.006). As a sign of bank solvency and stability, the average capital buffer in the sample is 0.11. The treatment group banks are more capitalized than control group banks (0.14 versus 0.10).

Table 3: **The effect of mandatory retention rule on banks' risk shifting**

	Dependent variable: RWA ratio				
	(1)	(2)	(3)	(4)	(5)
	RMBS≥0.5	RMBS≥0.8	RMBS≥0.8	RMBS=1	RMBS=1
					>= $250 billion included
Treatment	0.023*	0.023*	0.021*	0.024*	0.020*
	(0.013)	(0.013)	(0.011)	(0.012)	(0.011)
Capital buffer			0.001*	0.001*	-0.000
			(0.000)	(0.000)	(0.000)
Size			-0.097*	-0.085	0.020
			(0.049)	(0.053)	(0.040)
ROA			0.720**	0.625*	0.728**
			(0.332)	(0.342)	(0.294)
Liquidity ratio			-0.580***	-0.607***	-0.688***
			(0.136)	(0.119)	(0.118)
Securitization-assets ratio			0.077***	0.069**	0.071***
			(0.026)	(0.028)	(0.024)
Observations	333	325	325	293	325
R-squared	0.083	0.084	0.221	0.208	0.478
Number of banks	49	48	48	45	50

Notes: This table reports the effects of mandatory retention rules on banks' risk-weighted assets (RWA) ratio, specified by

$$RWA_{it} = ZD_{it} + X_{it}'\delta + \alpha_i + \gamma_t + \epsilon_{it}.$$

Treatment equals 1 if bank i is an RMBS securitizer after December 2015 and 0 otherwise. In the first 4 columns, I increase the threshold of RMBS activity share from 0.5 to 1. In the first two columns, only bank and time fixed effects are considered. In columns 3 and 4, bank-specific control variables are added. Banks with total assets of more than $250 billion are removed in the first 4 columns and added in column 5. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

Table 3 reports the estimation of equation (2). The first two columns report the results when no other control variables are included. The treatment groups are defined as RMBS activity share $\geq$ 0.5 and 0.8. The RWA ratio of the treatment banks increases 2.3 percentage points on average and is significantly positive at the 10% level. I increase the activity share threshold to 1 in column 4, and there is almost no discrepancy between the treatment effect estimates. As a robustness check, I add back banks with total assets greater than $250 billion and obtain a similar result, reported in column 5. I report in Table 8 in Appendix C that coefficients are similar when the dependent variable is replaced with the RWA ratio excluding securitization exposure. Hence, bank portfolio risk increased outside retained tranches in securitization.

Table 4: **The dynamics of treatment effects**

	Dependent variable: RWA ratio		
	RMBS $\geq$ 0.5	RMBS $\geq$ 0.8	RMBS = 1
Before4	0.003	0.003	0.001
	(0.016)	(0.016)	(0.016)
Before3	0.001	-0.001	-0.004
	(0.008)	(0.008)	(0.008)
Before2	0.010	0.007	0.004
	(0.010)	(0.010)	(0.010)
After0	0.021*	0.021*	0.021*
	(0.012)	(0.012)	(0.012)
After1	0.028**	0.027**	0.025**
	(0.012)	(0.012)	(0.012)
After2	0.022	0.022	0.022
	(0.014)	(0.014)	(0.015)
After3	0.034**	0.034**	0.033**
	(0.014)	(0.014)	(0.014)
Observations	333	325	293
R-squared	0.091	0.092	0.100
Number of banks	49	48	45

Notes: This table reports the dynamics of average treatment effects estimated by

$$RWA_{it} = \alpha_i + \sum_{N=2}^{4} Z_{5-N} B_{N,it} + \sum_{M=0}^{3} \alpha_{3-M} A_{M,it} + X'_{it}\delta + C_t + \epsilon_{it},$$

where B_N denotes N periods before the treatment and A_M denotes M periods after treatment. The quarter before treatment is used as the benchmark. Banks with total assets of more than $250 billion are removed. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

In Table 4, I report the estimation of equation (3) regarding the dynamics of banks' risk shifting with bank-specific controls. The coefficients on $B_{N,it}$ are all insignificant across all treatment groups, some of them being negative, indicating that prior to the Final Rule's implementation there is no difference in risk shifting measured by the RWA ratio; hence, there are no pre-existing trends. I find positive and significant coefficients on A_g: the results in the previous table occur after the Final Rule's implementation. Figure 7 plots the treatment effect coefficients and graphically illustrates the observed time pattern of the RWA ratio around the implementation of the mandatory retention rule. Each point on the graph reflects the average difference in the RWA ratio for treatment BHCs and control BHCs, netting out bank and time fixed effects, where the treatment group consists of all RMBS-only BHC securitizers. I use the quarter before the treatment (before1) as a base, represented by the dashed line. Bands represent 1.96 standard error of each point estimate.

Figure 7: Time passage relative to the mandatory retention rule's implementation

It's important to note that banks have different shares of RMBS securitization activities, resulting in varying levels of treatment intensity. Hence, in the continuous treatment study, I define the continuous treatment as the product of RMBS share and a time dummy that equals 1 if time is after December 2015. The econometric specification is

$$RWA_{it} = ZShare_{it} * D_t + X'_{it}\delta + \alpha_i + \gamma_t + \epsilon_{it}, \quad (4)$$

The coefficient of interest is Z. Banks' specific controls remain the same.

Table 5 displays the regression results of equation (4). The coefficient is significant across all specifications and roughly 0.02 in magnitude. This suggests that banks with a 10 percentage point higher share in RMBS activities are expected to increase their RWA ratio by 0.2 percentage points after the implementation of the retention rules. Notably, the coefficient does not vary significantly between the binary and continuous treatment, as the distribution of RMBS activity shares exhibits two mass points at 0 and 1.

Lastly, I also use the above DID design to study the impact of the retention rules on bank delinquency rates. The delinquency rate is defined in the same manner as in the motivating facts section, and specific details are presented in Tables 9 and 10 in Appendix C. Across all specifications, the results show that the delinquency rate decreased by approximately 0.3 percentage points. However, a concern with this approach is that delinquency typically occurs at least 90 days after a loan is generated, which raises doubts about the accuracy of the estimates derived from a short sample period. Consequently, I explore an alternative DID setup in the following section.

Table 5: **The effect of the mandatory retention rule on banks' risk shifting**

	Dependent variable: RWA ratio		
	(1)	(2)	(3)
Treatment	0.022*	0.020**	0.018*
	(0.011)	(0.010)	(0.010)
On-balance-sheet items	No	Yes	Yes
Off-balance-sheet items	No	No	Yes
Observations	359	359	359
R-squared	0.085	0.449	0.473
Number of bank	52	52	52

Notes: This table reports the effects of mandatory retention rules on banks' risk-weighted assets (RWA) ratio using continuous treatment specification

$$RWA_{it} = ZShare_{it} * D_t + X'_{it}\delta + \alpha_i + \gamma_t + \epsilon_{it}.$$

Treatment is the product of RMBS share and a dummy that equals 1 if time is after December 2015 and 0 otherwise. In the first column, only bank and time fixed effects are considered. In columns 2 and 3, bank-specific control variables are added. Banks with total assets more than $250 billion are removed. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

4.2 DID: RMBS-only securitizers versus mortgage sellers

This section employs an alternative DID method to study the impact of the retention rules on banks' delinquency rates. The analysis is based on a new sample range spanning from 2014Q1 to 2017Q4. The treatment group comprises RMBS-only BHC securitizers, i.e., those with an activity share equal to 1 in the previous DID setup, while the control group consists of non-securitizers that sell mortgages. The focus is on RMBS-only securitizers to avoid double impact of mandatory retention rules on some banks. I define a BHC as a mortgage seller if it has at least one quarter of non-zero mortgage sales. The resulting analysis includes 35 banks in the treatment group and 194 banks in the control group.

Table 6: **Summary statistics**

	Control group		Treatment group	
Variables	Mean	Std Dev	Mean	Std Dev
Total assets ($ billions)	6.579	16.649	24.738	36.819
DEL rate	0.021	0.027	0.010	0.015
Loan ratio (of total assets)	0.701	0.089	0.690	0.127
Home mortgage (of total loans)	0.284	0.165	0.280	0.165
Consumer loans (of total loans)	0.080	0.100	0.048	0.072
C&I loans (of total loans)	0.192	0.109	0.156	0.099
Capital buffer	0.111	0.667	0.073	0.040
Deposits ratio	0.619	0.111	0.658	0.114
Core deposits ratio	0.565	0.114	0.523	0.078
Liquidity ratio	0.218	0.080	0.250	0.122
ROA	0.006	0.005	0.006	0.005
Number of banks	194		35	

Notes: The table describes the control group and treatment group from 2014Q1 to 2017Q4. The treatment group consists of RMBS-only BHC securitizers (corresponding to RMBS securitizers with activity share equal to 1 in the previous DID setup), and the control group contains non-securitizers who are mortgage sellers. Mega banks (assets greater than $250 billions) are also excluded.

Table 6 presents the summary statistics. The treatment group consists of much larger banks, with assets valued at $25 billion, compared to $7 billion for the control group. However, both groups have similar asset compositions, with 70% of total assets in the form of loans and 20% in liquid assets. The loan types are also comparable across the two groups. Furthermore, the two groups exhibit the same level of profitability at 0.006. On the liability side, the treatment group is more well-capitalized with a capital buffer of 11%, whereas the control group has a capital buffer of 7%. Additionally, the treatment group borrows less from the deposit market, at 62%, compared to 66% for the control group.

In the benchmark, I consider the following empirical specification

$$DEL_{it} = \alpha_i + ZD_{it} + Y_{it}'\delta + C_t + \epsilon_{it}, \tag{5}$$

where DEL_{it} is the delinquency rate. Variable D_{it} equals 1 if bank i is an RMBS-only securitizer after December 2015 and 0 otherwise. Given the large sample size, I then employ propensity score matching (PSM) to reduce baseline bias. To calculate the propensity score, I select a set of observable variables for the logit group predicting function. Each bank is assigned a score, representing their probability of being in the treatment group. Banks in the treatment group receive a weight of 1, while banks in the control group receive a weight proportional to their probability of being in the treatment group relative to their probability of being in the control group.

Using the matched sample, I estimate the equation:

$$\Delta DEL_i = ZT_i + \epsilon_i \tag{6}$$

. Here DEL_i is the delinquency rate, and ΔDEL_i is defined as

$$\Delta DEL_i = \overline{DEL}_{i,after} - \overline{DEL}_{i,before},$$

which is the difference between the individual mean delinquency rates before and after December 2015, and T_i equals 1 if bank i is in the control group.

The set of covariates in the matching includes the common ones already mentioned: size, capital buffer, and ROA (profitability). Large banks tend to engage in more securitization activities. Securitization choice is also closely related to bank profitability, how tight of a capital constraint the bank faces. In Appendix C, I illustrate in Figure 15 the matching efficiency: selection bias (in terms of measured and tested covariates) in size, and ROA is reduced by matching. It also portrays the kernel distribution of propensity scores when banks are matched according to the common covariates above.

Table 7: **The effects of the mandatory retention rule on bank risk**

	Dependent variable: Delinquency rate		
	(1)	(2)	(3)
	OLS	OLS	PSM
Treatment	-0.003	-0.003*	-0.002**
	(0.002)	(0.002)	(0.001)
Bank controls	No	Yes	
Number of banks	229	228	220

Notes: This table reports the effects of mandatory retention rules on banks' RWA ratios and delinquency rates under propensity score matching. Columns 1 and 2 are the benchmark cases without matching. Columns 3 and 4 report the PSMDID estimates using different sets of covariates. *Treatment* equals 1 if bank i is an RMBS-only securitizer after December 2015 and 0 otherwise. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

Table 7 reports the results. There is a decline in the delinquency rate with and without matching techniques. In OLS, the drop is significant when bank controls are added. Under matching, the drop is 0.2 percentage points and significant at the 5% level using common covariates. I also portray the dynamics of the treatment effects in Figure 8. The results take effect after the effective date of the Dodd-Frank retention rules in RMBS.

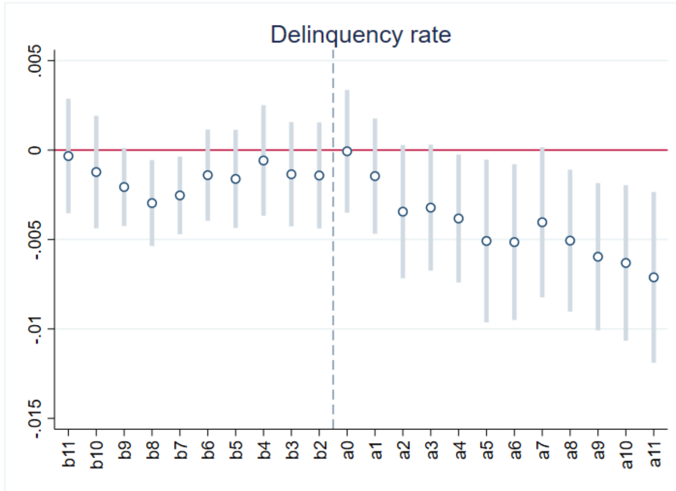

Figure 8: Time passage relative to the mandatory retention rule's implementation (from 2013Q1 to 2018Q4)

Lastly, I also use the PSM DID methodology to study the impact of retention rules on the RWA ratio. Here I introduce an additional covariate that is related to banks' incentive for risk shiftng: the core deposits ratio, which measures the fraction of stable sources of funding. The results are reported in Table 11 in Appendix C. In the OLS regression, the treatment coefficient is not significant and is inconsistent depending on whether bank-specific controls are added or not. The substantial difference in size between the two groups makes them incomparable. However, under matching, the coefficient becomes significant at the 10% level. The increase in the RWA ratio is 2.1 percentage points, similar to the finding in the previous analysis.

In sum, the above two sections show that loans are riskier ex ante and safer ex post. Consequently, banks shifted toward riskier investing strategies and must have engaged in more efforts to monitor their borrowers. According to the model prediction, the current 5% ratio is too high relative to the optimal retention level.

5 Extension: Optimal Retention Form

The model in Section 3 assumes that the bank retains the vertical interest. This section studies three questions: 1) what are the bank's behaviors under the horizontal retention, 2) which retention form is optimal from the bank's perspective, and 3) which retention form is optimal from the welfare perspective. They are discussed in the subsequent three subsections.

5.1 Horizontal retention

In a horizontal retention, the bank retains Z fraction of the most junior tranches in the pool. Recall that r is the realization in the bad state. In this section, assume $r \geq 0$. As a starting point, the market value of the pool given M can be rewritten as

$$\alpha R - \big(1 - q_K(M)\big)\frac{R-r}{R}\alpha R,$$

which is the promised payment net of expected loss. The realization of r means a loss, and $\frac{R-r}{R}$ measures the loss severity as the percentage lost in the event of default. The default risk of the underlying project is then disproportionally distributed between investors and the bank. By seniority, investors start to bear loss only if the loss severity $\frac{R-r}{R}$ is above Z, or

$$Z \leq \frac{R-r}{R}. \tag{7}$$

The analysis of monitoring choice at time 2 remains the same: the bank chooses $M = 1$ if $Z \geq \bar{Z}_K$. We make the following assumption that r is not large.

Assumption 5 $r < R_g - d - \frac{c}{\Delta_g}$.

Under this assumption, the loss severity $\frac{R-r}{R}$ is always larger than $\bar{Z}$, where $\bar{Z}$ is defined in Section 3.2.

The price schedule investors provide is the following

$$P^h(Z|K) = \begin{cases} (1-Z)\alpha R & \text{if } Z \geq \frac{R-r}{R}, \\ (1-Z)\alpha R - (1-p)\alpha\big(\frac{R-r}{R} - Z\big)R & \text{if } \bar{Z}_K \leq Z \leq \frac{R-r}{R}, \\ (1-Z)\alpha R - (1-p+\Delta)\alpha\big(\frac{R-r}{R} - Z\big)R & \text{if } Z < \bar{Z}_K. \end{cases}$$

The bank's monitoring decision affects the probability of investors receiving losses, which is $1-q_K(M)$. But they assume losses only amount to $\frac{R-r}{R} - Z$. Therefore when $Z \geq \frac{R-r}{R}$, investors receive the full value of senior tranches that are worth $(1-Z)\alpha R$. Let $\bar{R} \equiv pR + (1-p)r$ and $\underline{R} \equiv (p-\Delta)R + (1-p+\Delta)r$ denote the expected returns of the project under monitoring and no monitoring. The value of the retained tranches, denoted by $V^h(Z|K)$, is the total value of the pool minus the value of the tranches sold: $V^h(Z|K) = V^{\text{pool}} - P^h(Z|K)$. When $Z \geq \bar{Z}_K$, the pool value is $V^{\text{pool}} = \alpha\bar{R}$; and when Z is below $\bar{Z}_K$, the pool value drops to $V^{\text{pool}} = \alpha\underline{R}$. Hence,

$$V^h(Z|K) = \begin{cases} \alpha(p+Z-1)R + \alpha(1-p)r & \text{if } Z \geq \frac{R-r}{R} \\ pZ\alpha R & \text{if } \bar{Z}_K \leq Z \leq \frac{R-r}{R} \\ (p-\Delta)Z\alpha R & \text{if } Z < \bar{Z}_K. \end{cases}$$

And the bank's retention constraint, $\frac{V^h(Z|K)}{V^{\text{pool}}} \geq \theta$, can be written as

$$\begin{cases} ZR \geq \theta\bar{R} + (1-p)(R-r) & \text{if } Z \geq \frac{R-r}{R} \\ ZpR \geq \theta\bar{R} & \text{if } \bar{Z}_K \leq Z \leq \frac{R-r}{R} \\ Z(p-\Delta)R \geq \theta\underline{R} & \text{if } Z < \bar{Z}_K. \end{cases}$$

For any given θ, the retention constraint imposes a lower bound for Z that is strictly larger than θ, due to the disproportional risk sharing under the horizontal retention. When risk is mostly absorbed by the junior tranches, the value of the tranches is discounted, or smaller than θ fraction of the pool value. The bank thus has to retain more to meet the value requirement. One result follows from the above expressions.

Proposition 5.1 *If $r = 0$, the vertical and horizontal retentions are identical.*

When $r = 0$, investors cannot receive the full value of the tranches under any circumstances. When $r > 0$, the bank's objective function at time 1, $\lambda P^h(Z|K) + \Pi_2(Z|K)$, has the general shape in Figure 9. The objective function decreases on $[\bar{Z}, 1]$ but decreases faster when $Z \geq \frac{R-r}{R}$ because of the depletion of the junior tranches in this region.

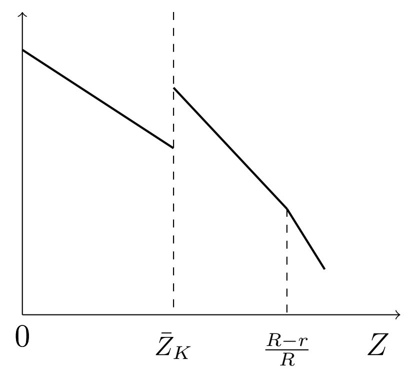

Figure 9: The bank's objective function

Let $\underline{Z}'_K$ be such that $\lambda P^h(\underline{Z}'_K|K) + \Pi_2(\underline{Z}'_K|K) = \lambda P^h(\bar{Z}_K|K) + \Pi_2(\bar{Z}_K|K)$, similar to $\underline{Z}_K$ in the vertical retention. The solution to the bank's retention problem is the following:

Proposition 5.2 (Retention choice.) *Under horizontal retention, the bank's optimal retention choice of Z is*

$$Z^h(\theta) = \begin{cases} \frac{\theta\underline{R}}{(p-\Delta)R} & \text{if } \theta < \frac{\underline{Z}'_K(p-\Delta)R}{\underline{R}_K} \\ \bar{Z}_K & \text{if } \frac{\underline{Z}'_K(p-\Delta)R}{\underline{R}} \leq \theta < \frac{\bar{Z}_K pR}{\bar{R}} \\ \frac{\theta\bar{R}}{pR} & \text{if } \frac{\bar{Z}_K pR}{\bar{R}} \leq \theta \leq \frac{R-r}{p\bar{R}} \\ \frac{\theta\bar{R}+(1-p)(R-r)}{R} & \text{if } \theta \geq \frac{R-r}{p\bar{R}} \end{cases}$$

Moreover, $Z^h(\theta) \geq \theta$. The equality holds only at 0 and 1.

Again, due to seniority, the junior tranches bear most of the risk, which discounts their value. The bank has to retain a higher fraction of the pool to meet the requirement. The bank's expected value, denoted by

$$\Pi_1(\theta|K) \equiv \lambda P^h(Z^h(\theta)|K) + \Pi_2(Z^h(\theta)|K),$$

is portrayed in Figure 10. Besides the positions of the kink points, the only difference between this value under horizontal retention and vertical retention occurs for very large θ. Similar analysis can be conducted for the project choice stage, although the more retention leading to more risk shifting argument is omitted here.

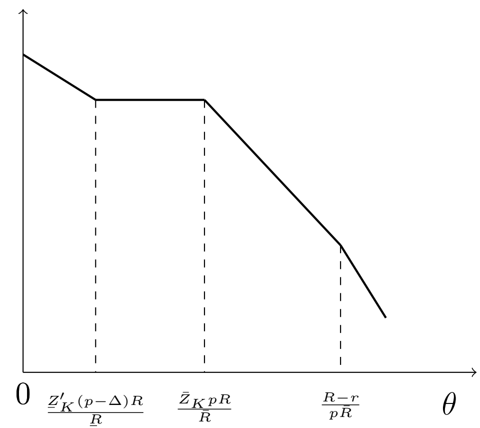

Figure 10: The bank's expected value under an optimal horizontal retention choice

5.2 The banks' optimal choice of retention form

In practice, with horizontal retention, higher risk weights will be attached to the junior tranches retained by the bank. This can result in the bank needing to issue additional equity to meet capital requirements, which can be costly. Based on the results above, I argue here that given a project K, if capital requirements are not a concern, or "slack" for the bank, then the expected value under horizontal retention is higher than under vertical retention.

Proposition 5.3 *If the capital requirement is slack at retention requirement θ, the bank chooses horizontal retention.*

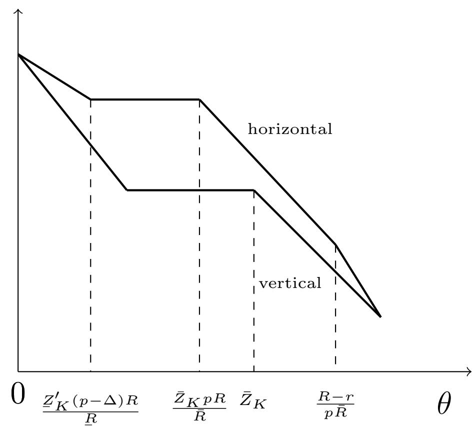

Figure 11: The bank's expected value under horizontal and vertical retention choices

Horizontal retention necessitates that banks retain more on their balance sheets because of the increased risk involved, resulting in greater gains in the good states and larger losses during defaults. As the bank already loses everything in bankruptcy due to limited liability, the net expected gains are higher. This result is better illustrated in Figure 11.

If the bank is unable to fulfill the capital requirement by holding a horizontal piece, and the cost of issuing additional equity is too high, then the bank must opt for the vertical form. In practice, banks are heterogeneous and retain differently. For example, Flynn, Ghent, and Tchistyi (2020) examine CMBS deals from 2017 to 2019 and show that 45% of the deals involve horizontal retention and 40% involve vertical retention (though the issuers may not be bank holding companies). Other reasons to hold a vertical piece include consolidation concerns[18] and opportunity cost, because senior tranches can be collateral to borrow repo and other short-term funding. The discussion above hence predicts that banks that retain vertically should perform better and accumulate less risk.

5.3 Optimal retention form

The remaining question is, what is the optimal retention form from the welfare perspective? In fact, policymakers expect that incentives are better aligned under horizontal risk retention.

[18]There can be the potential impact of accounting standards FAS 166 and FAS 167. Effective as of 2010, the two accounting standards tightened the accounting for securitizations and the consolidation of variable interest entities (VIEs). If a securitizer is identified as the primary beneficiary for the special-purpose entity used to issue ABS, it must consolidate the securitized assets. The 5% retention may meet the threshold of an interest that could potentially be significant. Furthermore, the horizontal strip of risk is more likely to be significant from the results in the last section. If the consolidation cost is high enough, retaining a horizontal component is no longer optimal if holding a vertical strip can avoid such consolidation.

For example, the final rules (p. 486) state

> "*...a sponsor choosing to retain risk in a more expensive horizontal form over a vertical form would have greater exposure to credit risk, and that sponsor's incentives should be better aligned with investors.*"

This is true with regard to monitoring, according to the observations in the previous section. Recall that the welfare-maximizing retention ratio is $[\underline{Z}_g, \bar{Z}_g]$ under vertical retention and $[\frac{\underline{Z}'_g(1-\Delta_g)R_g}{\underline{R}_g}, \frac{\bar{Z}_g p_g R_g}{\bar{R}_g}]$ under horizontal retention. Since $\frac{\underline{Z}'_g(1-\Delta_g)R_g}{\underline{R}_g} < \underline{Z}_g$ and $\frac{\bar{Z}_g R_g}{\bar{R}_g} < \bar{Z}_g$, as shown in Figure 11, it is easier to induce monitoring under horizontal retention but also easier to over-regulate securitization. Moreover, concerning risk shifting, the threshold above which banks start to gamble is also lower under horizontal retention, implying that banks are more likely to shift risk under horizontal retention. The reason is that the horizontal piece requires the bank to retain a higher fraction of the pool, which resembles an increased retention ratio, under which gambling is more attractive. Therefore, there is also a trade-off between the two retention forms. This is in the spirit of Fender and Mitchell (2009). But due to the costly equity issuance and accounting issues about consolidation, policymakers should give discretion to banks.

6 Concluding Remarks

Policymakers have implemented new risk retention rules with the expectation of aligning incentives and reducing banks' risk. However, I have demonstrated that these rules involve trade-offs among various dimensions of bank risk taking. While mandatory retention may result in more good screening and monitoring, it may also lead to a riskier portfolio. In the US BHC data, I have observed that banks' portfolios became riskier ex ante but had fewer delinquencies ex post after the implementation of the new rules. Importantly, my theory indicates that the simultaneous increase in monitoring and risk shifting occurs only when the retention ratio exceeds the optimal level. This suggests that the current 5% ratio is too high. While the optimal level is an interior number, a structural model is needed to precisely quantify it, and this is left for future research.

References

Agarwal, Sumit, Brent W Ambrose, Yildiray Yildirim, and Jian Zhang. 2019. "Risk retention and qualified commercial mortgages." *Working Paper, National University of Singapore* .

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff. 2011. "The role of securitization in mortgage renegotiation." *Journal of Financial Economics* 102:559–578.

Chemla, Gilles and Christopher A. Hennessy. 2014. "Skin in the game and moral hazard." *The Journal of Finance* 69 (4):1597–1641.

Chen, Weitzu, Chi-Chun Liu, and Stephen G. Ryan. 2008. "Characteristics of Securitizations That Determine Issuers' Retention of the Risks of the Securitized Assets." *The Accounting Review* 83 (5):1181–1215.

DeMarzo, Peter and Darrell Duffie. 1999. "A Liquidity-Based Model of Security Design." *Econometrica* 67 (1):65–100.

Demsetz, Rebecca S., Marc R. Saidenberg, and Philip E. Strahan. 1996. "Banks with something to lose: The disciplinary role of franchise value." *FRBNY Economic Policy Review* 2 (2).

Erel, Isil, Taylor Nadauld, and René M. Stulz. 2013. "Why Did Holdings of Highly Rated Securitization Tranches Differ So Much across Banks?" *The Review of Financial Studies* 27 (2):404–453.

Fender, Ingo and Janet Mitchell. 2009. "Incentives and tranche retention in securitisation: a screening model." *Working paper, Centre for Economic Policy Research* .

Flynn, Jr., Sean J, Andra C Ghent, and Alexei Tchistyi. 2020. "Informational efficiency in securitization after Dodd-Frank." *The Review of Financial Studies* 33 (11):5131–5172.

Furfine, Craig. 2020. "The Impact of Risk Retention Regulation on the Underwriting of Securitized Mortgages." *Journal of Financial Services Research* 58 (2):91–114.

Furlong, Frederick T. 1988. "Changes in Bank Risk-taking." *Federal Reserve Bank of San Francisco Economic Review* 2:45–56.

Gorton, Gary and George Pennacchi. 1995. "Banks and loan sales Marketing nonmarketable assets." *Journal of Monetary Economics* 35 (3):389–4110.

Gorton, Gary and Andrew Winton. 2003. "Financial intermediation." In *Handbook of the Economics of Finance*, vol. 1, Part 1, edited by G.M. Constantinides, M. Harris, and R. M. Stulz, chap. 08. Elsevier, 1 ed., 431–552.

Guo, Guixia and Ho-Mou Wu. 2014. "A study on risk retention regulation in asset securitization process." *Journal of Banking and Finance* 45 (C):61–71.

Jensen, Michael and William H Meckling. 1976. "Theory of the firm: Managerial behavior, agency costs and ownership structure." *Journal of Financial Economics* 3 (4):305–360.

Keeley, Michael C. 1990. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80 (5):1183–1200.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did securitization lead to lax screening? Evidence from subprime loans." *The Quarterly Journal of Economics* 125 (1):307–362.

Kiff, John and Michael Kisser. 2014. "A shot at regulating securitization." *Journal of Financial Stability* 10 (C):32–49.

Leland, Hayne and David H Pyle. 1977. "Informational Asymmetries, Financial Structure, and Financial Intermediation." *Journal of Finance* 32 (2):371–87.

Malekan, Sara and Georges Dionne. 2014. "Securitization and optimal retention under moral hazard." *Journal of Mathematical Economics* 55 (C):74–85.

Mian, Atif and Amir Sufi. 2009. "The consequences of mortgage credit expansion: evidence from the U.S. mortgage default crisis." *The Quarterly Journal of Economics* 124 (4):1449–1496.

Mitchell, Janet. 2004. "Financial intermediation theory and the sources of value in structured finance markets." *Working Paper No. 71, National Bank of Belgium* .

Pagès, Henri. 2013. "Bank monitoring incentives and optimal ABS." *Journal of Financial Intermediation* 22 (1):30–54.

Parlour, Christine A. and Guillaume Plantin. 2008. "Loan sales and relationship banking." *The Journal of Finance* 63 (3):1291–1314.

Pennacchi, George. 1988. "Loan sales and the cost of bank capital." *The Journal of Finance* 43 (2):375–396.

Petersen, Mitchell A. and Raghuram G. Rajan. 2002. "Does distance still matter? The information revolution in small business lending." *The Journal of Finance* 57 (6):2533–2570.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig. 2010. "Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis." *Journal of Financial Economics* 97:369–397.

Purnanandam, Amiyatosh. 2011. "Originate-to-distribute Model and the Subprime Mortgage Crisis." *The Review of Financial Studies* 24 (6):1881–1915.

Repullo, Rafael. 2004. "Capital requirements, market power, and risk-taking in banking." *Journal of Financial Intermediation* 13 (2):156–182.

Sarkisyan, Anna and Barbara Casu. 2013. "Retained interests in securitizations and implications for bank solvency." *Working Paper Series 1538, European Central Bank* .

Appendix A Proofs

A.1 Proof of Proposition 3.1

Given (K, Z), the bank monitors if

$$p_K\big((1-\alpha)R_K + R_K - d\big) - c \geq (p_K - \Delta_K)\big((1-\alpha)R_K + R_K - d\big).$$

The result follows. □

A.2 Proof of Proposition 3.2

For $Z < \bar{Z}$, the bank does not monitor. The bank's objective function is

$$\lambda(1-Z)(p-\Delta)R + (p-\Delta)\Bigg((1-\alpha)R + Z\alpha R - d\Bigg)$$

and is linear and decreasing since $\lambda > 1$. For $Z \geq \bar{Z}$, the bank monitors, and its objective function is

$$\lambda(1-Z)\alpha pR + p\Bigg((1-\alpha)R + Z\alpha R - d\Bigg) - c$$

and is linear and decreasing by the same token. To show the upper semicontinuity, note that $\Pi_2(Z|K)$ is continuous at $\bar{Z}$, and

$$(1-\bar{Z})\alpha pR > (1-\bar{Z})\alpha(p-\Delta)R$$

hence the value of the objective function at $\bar{Z}$ is strictly larger than its left limit. □

A.3 Proof of Proposition 3.4

We only need to show the second part of the result. It suffices to show $\underline{Z}_g > \underline{Z}_b$, which comes from the following comparisons

$$\begin{aligned} p_b(R_b - d) &> (R_g - d), \\ R_b &> R_g, \\ \frac{c}{\Delta_g} &< \frac{c}{\Delta_b}. \end{aligned}$$

A.4 Proof of Proposition 3.5

The proof is divided into three steps.

Step 1. In this step, we show that $G(\theta)$ is increasing in θ when the retention constraints are binding for both projects; that is, when θ is below $\underline{Z}_b$ and when θ is above $\bar{Z}_g$. In the first region, there is no monitoring,

$$\begin{aligned} G(\theta) =& \lambda(1-\theta)\alpha((p_b-\Delta_b)R_b-(1-\Delta_g)R_g) \\ &+ (p_b-\Delta_b)\big((1-\alpha)R_b+\theta\alpha R_b-d\big)-(1-\Delta_g)\big((1-\alpha)R_g+\theta\alpha R_g-d\big) \end{aligned}$$

It follows that $G'(\theta)=(\lambda-1)\alpha((1-\Delta_g)R_g-(p_b-\Delta_b)R_b)>0$. In the second region, the bank monitors for both projects, and

$$G(\theta) = \lambda(1-\theta)\alpha(p_bR_b-R_g)+p_b\big((1-\alpha)R_b+\theta\alpha R_b-d\big)-\big((1-\alpha)R_g+\theta\alpha R_g-d\big).$$

By the same token, $G'(\theta)=(\lambda-1)\alpha(R_g-p_bR_b)>0$.

Step 2. In this step, we pin down $\bar{\theta}$ and show that $G(\theta)$ is positive if and only if θ is above $\bar{\theta}$. In fact, $\bar{\theta}$ is such that

$$G(\bar{\theta})=0$$

when θ is above $\bar{Z}_g$. Specifically,

$$\bar{\theta}=\frac{(\lambda\alpha+1-\alpha)(R_g-p_bR_b)-(1-p_b)d}{(\lambda-1)\alpha(R_g-p_bR_b)}. \tag{8}$$

According to step 1, when θ is above $\bar{\theta}$, $G(\theta)$ is positive. We show that $G(\theta)<0$ when θ is below $\bar{\theta}$. There are two scenarios to consider. First, consider $\underline{Z}_g<\bar{Z}_b$. In this case, $G(\theta)$ is increasing on $[\underline{Z}_k,\underline{Z}_g]$ because $V_b(\theta)$ is constant, constant on $[\underline{Z}_g,\bar{Z}_b]$ because both V_g and V_b are constant, and decreasing on $[\bar{Z}_b,\bar{Z}_g]$ because $V_g(\theta)$ is constant. The maximum of $G(\theta)$ on this region is achieved on $[\underline{Z}_g,\bar{Z}_b]$. According to Assumption 3, this value is negative. Second, consider $\underline{Z}_g>\bar{Z}_b$. In this case, $G(\theta)$ is increasing on $[\underline{Z}_k,\bar{Z}_b]$ because $V_b(\theta)$ is constant, and decreasing on $[\underline{Z}_g,\bar{Z}_g]$ because $V_g(\theta)$ is constant. On $[\underline{Z}_k,\underline{Z}_g]$ $G(\theta)$ can be increasing or decreasing, depending on the comparison of p_bR_b and $(1-\Delta_g)R_g$, because the bank monitors the gambling project but not the good one. We will assume later that $p_b>1-\Delta_g$, hence $p_bR_b>(1-\Delta_g)R_g$, and the maximum is achieved at $\bar{Z}_b$. At this point,

$$V_b(\bar{Z}_b)-V_g(\bar{Z}_b)<V_b(\bar{Z}_b)-V_g(\bar{Z}_g)<0.$$

This step is completed.

Step 3. Let $\underline{\theta}\equiv\underline{Z}_g$. Assumption 3 also guarantees that $\underline{\theta}<\bar{\theta}$. When $\theta\in[\underline{\theta},\bar{\theta}]$, the bank chooses the good project and monitors. The proof is completed. □

A.5 Proof of Proposition 3.6

One needs to compare the welfare at $\theta=0$ and $\underline{Z}_g$. The latter is higher under the above assumption. □

A.6 Proof of Proposition 5.1

If $r = 0$, $\bar{R} = pR$ and $\underline{R} = (1 - \Delta)R$. The rest follows. $\square$

A.7 Proof of Proposition 5.2

By the definition of $\underline{Z}'_K$, the retention constraint is not binding only when Z is between $\underline{Z}'_K$ and $\bar{Z}$. The results follow. $\square$

A.8 Proof of Proposition 5.3

When the retention constraint is binding under both retention forms, $P^h(Z^h(\theta)|K)$ is equal to $P^v(Z^v(\theta)|p, R)$, which are just $1 - \theta$ fraction of the total pool value, but $Z^h(\theta) > \theta = Z^v(\theta)$ at non-endpoints. Hence the retained tranches have higher expected value under horizontal retention. These together imply that a horizontal component generates higher expected value. This argument holds at $\bar{Z}$. With the fact that the value function is decreasing faster under vertical retention for θ below $\frac{R-r}{pR}$, and the two value functions coincide at 0 and 1, the results follow. $\square$

Appendix B Ex ante screening

In this section, I present a different version of the model that involves ex ante screening.

B.1 Setup

Prior to a securitization transaction, a bank has the ability to manipulate the quality of its asset pool by exerting a screening effort, denoted as $s \in \{0, 1\}$. The screening technology is the same as the monitoring technology in the main model:

$$p(s) = p - (1 - s)\Delta,$$

with $\Delta_b > \Delta_g$, and the cost of screening is c. The screening effort is unobservable. However, investors are assumed to be sophisticated enough to take into account the bank's incentive problem. This distinguishes the model from signaling models where securitizers signal asset quality by selecting different retention fractions or forms (e.g., see DeMarzo and Duffie (1999) and Flynn, Ghent, and Tchistyi (2020)). Investors' belief about the level of the bank's screening effort is denoted as s'. To maintain consistency with the original model, I assume that $r = 0$, which implies that the two retention forms are identical and the retention constraint is simply $Z \geq \theta$. I also reorder the dates to $t \in \{-1, 0, 1, 2\}$. The timeline of events is

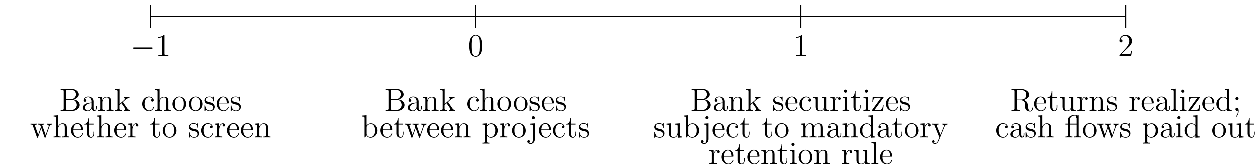

Figure 12: Timeline of the events

The equilibrium definition in the sense of subgame perfect equilibrium is the following.

Definition B.1 (Equilibrium.) *The equilibrium consists of the bank's screening choice s at time -1; the bank's project choice K at time 0; the bank's retention choice Z at time 1; prices P of the tranches sold to investors and investors' belief on screening effort s' such that P reflects project choice and retention choice, and investors' belief; investors break even; the bank maximizes its value sequentially subject to the retention constraint at time 1; investors' belief is correct: $s' = s$.*

B.2 Equilibrium

At time 1, the price schedule offered by investors is

$$P(Z, p, R, s') = (1 - Z)\alpha p(s')R$$

The bank hence solves at time 1

$$\Pi_1(K, s; \theta) = \max_{Z} \ \lambda P(Z, K, s') + p(s)\Big((1-\alpha)R + R - d\Big) - c(s)$$
$$\text{s.t. } Z \geq \theta$$

The bank's retention choice is the following.

Proposition B.1 (Retention choice under vertical component.) *If $s' \geq s$, the bank chooses $Z = \theta$. If $s' = 0$ and $s = 1$, the bank chooses*

$$Z = \begin{cases} \theta & \text{if } 1 < \frac{(\lambda-1)p}{\lambda\Delta} \\ 1 & \text{if } \frac{(\lambda-1)p}{\lambda\Delta} < 1 \end{cases}$$

Proof: The partial derivative with respect to Z is

$$-\alpha\big(\lambda p(s)R - p(s)R + \lambda(s' - s)\Delta R\big)$$

The result follows. □

In equilibrium, investors hold correct beliefs that the retention constraint will be binding, expressed as $Z = \theta$. By securitizing, the bank receives the full market value of the asset plus the securitization benefit, which overrides the expected project gain from retention.

At time 0, similar to the main model, the risk-shifting propensity is defined as

$$G(\theta, s) = \Pi_1(b, s; \theta) - \Pi_1(g, s; \theta).$$

The bank will choose to gamble if this gain is positive.

Proposition B.2 (More retention leads to more bank risk shifting.) *In equilibrium, $G(\theta, s)$ is increasing in θ for any s.*

Proof: In equilibrium, $s' = s$. The bank retains $Z = \theta$. We have

$$\frac{\partial G}{\partial \theta} = (\lambda - 1)(p_g(s)R_g - p_b(s)R_b) > 0$$

because $p_g(s)R_g > p_b(s)R_b$ for any s. □

The intuition behind this model is similar to the main model. At a low retention ratio, the bank selects the good project because of its favorable market value, which makes it easy to securitize. At a high retention ratio, the gambling project becomes more profitable due to limited liability and a higher upper side return R_b. Let $\bar{\theta}$ be such that $G(\bar{\theta}, 1) = 0$. In equilibrium, this threshold represents the point above which the bank shifts risk when it does not shirk in screening.

The next result presents a crowding-out theory of screening and risk shifting. I show that screening and choosing the good project can be substitutes.

Proposition B.3 (More screening leads to more bank risk shifting.) *In equilibrium, $G(\theta, 1) > G(\theta, 0)$.*

Proof: In equilibrium, we have

$$G(\theta, 1) - G(\theta, 0) = \Big(\lambda\alpha(1-\theta) + \alpha\theta + 1 - \alpha\Big)(\Delta_b R_b - \Delta_g R_g) > 0$$

which completes the proof. □

The return to screening, ΔR, is higher for the gambling project. Increasing screening effort leads to an increase in the upper side return of the project, and the expected value of the gambling project increases more than that of the good project. Screening and gambling are complements, and screening crowds out the good project in this sense.

Next, I examine the bank's screening motive and complete the characterization of the equilibrium. I have already shown that the risk-shifting propensity is increasing in screening effort, i.e., $G(\theta, 1)$ is greater than $G(\theta, 0)$. I have also defined $\bar{\theta}$ as the threshold satisfying $G(\bar{\theta}, 1) = 0$. Above this threshold, the bank always chooses to gamble if it has screened the pool of projects.

At the boundary $\theta = 1$, there are two possibilities for $G(\theta = 1, 0)$. I first discuss the scenario when $G\big(1, 0\big) \leq 0$, which means that gambling will not happen without screening. As As mentioned earlier, a lack of due diligence in screening causes the value of the gambling project to drop more, and this scenario is an extension of that result. The analysis is then divided into two parts: for $\theta \leq \bar{\theta}$ and for $\theta \in (\bar{\theta}, 1]$. In the first part, the bank always chooses the good project, due to the fact that

$$G(\theta, 0) < G(\theta, 1) < G(\bar{\theta}, 1) = 0.$$

That is, the gain from risk-shifting propensity is always negative, regardless of the bank's screening effort. Given the choice of the good project, the bank screens if the following condition is met

$$(1-\alpha)R_g + \alpha Z(\theta, 1, s')R_g - d - c \geq (1-\Delta_g)\big((1-\alpha)R_g + \alpha Z(\theta, 0, s')R_g - d\big) \tag{9}$$

The bank is comparing profits conditional on success. The payment from securitization purchasers is independent of the bank's screening effort. Therefore, it does not show up in the comparison. The bank's expected profits, or its skin in the game, are increasing in the retention ratio. In equilibrium, $Z = \theta$, the above condition (9) simplifies as

$$\theta \geq \underline{\theta} \equiv \frac{\frac{c}{\Delta_g} + d - (1-\alpha)R_g}{\alpha R_g}$$

This is the intended consequence of mandatory retention: the bank increases its effort in screening as the retention ratio rises. I summarize the equilibrium for $\theta \leq \bar{\theta}$.

Proposition B.4 *When $\theta < \underline{\theta}$, the bank chooses the good project but does not screen ex ante. When $\underline{\theta} \leq \theta \leq \bar{\theta}$, the bank chooses the good project and screens.*

Figure 13 portrays the risk-shifting propensities under different screening efforts on the θ space. Thanks to the results in the previous section, both propensities are increasing in θ, and $G(\theta, 1)$ is above $G(\theta, 0)$. Note that $G(\theta, 0)$ is steeper than $G(\theta, 1)$.

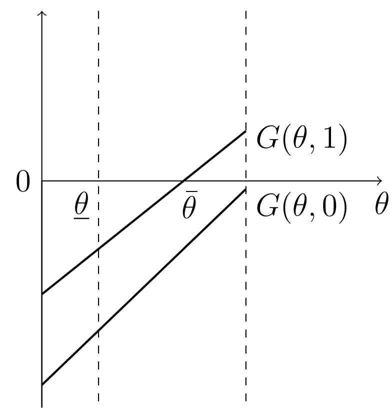

Figure 13: Risk-shifting propensity under screening choice

For θ greater than $\bar{\theta}$ but smaller than 1, the bank chooses the gambling project if there is screening and chooses the good project if there is no screening due to the fact that

$$G(\theta,0) < G\big(1,0\big) < 0 = G(\bar{\theta},1) < G(\theta,1)$$

It turns out that the bank will screen and gamble thereafter.

Proposition B.5 *When $\theta \in (\bar{\theta}, 1]$, the bank screens and gambles.*

Proof: Since $\theta > \underline{\theta}$, screening generates a higher value for the bank undertaking the good project. And under screening, the gambling project dominates the good project because $\theta > \bar{\theta}$. □

In sum, the equilibrium is similar to the one in the main model. The welfare analysis hence follows. The remaining case is when $G(1,0) > 0$, which is depicted in Figure 14.

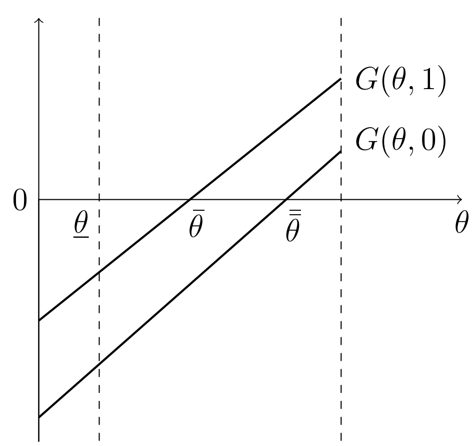

Figure 14: Risk-shifting propensity under screening choice

Let $\bar{\bar{\theta}}$ be such that $G(\bar{\bar{\theta}}, 0) = 0$. Then the analysis for $\theta \leq \bar{\bar{\theta}}$ is the same as the main model. The bank screens and gambles simultaneously on $(\bar{\theta}, \bar{\bar{\theta}}]$. On $(\bar{\bar{\theta}}, 1]$, the bank always gambles since both G functions are positive. It screens if

$$\theta \geq \frac{\frac{c}{\Delta_b} + d - (1 - \alpha)R_b}{\alpha R_b}$$

Since $R_b > R_g$, this threshold is smaller than $\underline{\theta}$, hence smaller than $\bar{\bar{\theta}}$, implying that the bank always screens on this interval. As a result, the threshold $\bar{\bar{\theta}}$ does not matter, and the bank screens and gambles on $(\bar{\theta}, 1]$. The equilibrium is the same.

Appendix C Supplementary tables and graphs

Table 8: **The effect of the mandatory retention rule on banks' risk shifting**

	Dependent variable: RWA ratio excluding securitization exposure			
	(1)	(2)	(3)	(4)
	RMBS≥0.5	RMBS≥0.8	RMBS=1	RMBS=1
Treatment	0.023*	0.023*	0.024*	0.023*
	(0.013)	(0.013)	(0.013)	(0.012)
Bank controls	No	No	No	Yes
Observations	331	323	291	291
R-squared	0.080	0.082	0.091	0.464
Number of banks	49	48	45	45

Notes: This table reports the effects of mandatory retention rules on banks' risk-weighted assets (RWA) ratio excluding securitization exposure, specified by

$$RWA_{it} = ZD_{it} + X_{it}'\delta + \alpha_i + \gamma_t + \epsilon_{it}$$

Treatment equals 1 if bank i is an RMBS securitizer after December 2015 and 0 otherwise. In the first 4 columns, I increase the threshold of RMBS activity share from 0.5 to 1. In the first two columns, only bank and time fixed effects are considered. In columns 3 and 4, bank-specific control variables are added. Banks with total assets more than $250 billion are removed in the first 4 columns and added in column 5. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

Table 9: **The effect of the retention rule on banks' loan delinquency rate**

	(1)	(2)	(3)
	RMBS≥0.5	RMBS=1	RMBS=1
Treatment	-0.003*	-0.003*	-0.004*
	(0.002)	(0.002)	(0.002)
Bank Controls	No	No	Yes
Observations	333	293	293
R-squared	0.089	0.107	0.161
Number of banks	49	45	45

Notes: This table reports the effects of mandatory retention rules on banks' loan delinquency ratio. The econometric specification is binary treatment

$$DEL_{it} = ZD_{it} + X'_{it}\delta + \alpha_i + \gamma_t + \epsilon_{it}$$

The sample period is from 2015Q1 to 2016Q4. *Treatment* equals 1 if bank i is an RMBS securitizer after December 2015 and 0 otherwise. In the first 2 columns, I increase the threshold of RMBS activity share from 0.5 to 1. In the first two columns, only bank and time fixed effects are considered. In column 3, bank-specific control variables are added. Banks with total assets more than \$250 billion are removed. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

Table 10: **The effect of the retention rules on banks' loan delinquency rate (continuous treatment)**

	(1)	(2)	(3)
Treatment	-0.004*	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)
Bank Controls	No	No	Yes
Observations	333	293	293
R-squared	0.089	0.107	0.161
Number of bank	49	45	45

Notes: This table reports the effects of mandatory retention rules on banks' loan delinquency ratio. The econometric specification is continuous treatment

$$DEL_{it} = ZShare_{it} * D_t + X'_{it}\delta + \alpha_i + \gamma_t + \epsilon_{it}$$

The sample period is from 2015Q1 to 2016Q4. D_t equals 1 if time t is after December 2015 and 0 otherwise. In the first 2 columns, I increase the threshold of RMBS activity share from 0.5 to 1. In the first two columns, only bank and time fixed effects are considered. In column 3, bank-specific control variables are added. Banks with total assets more than \$250 billion are removed. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

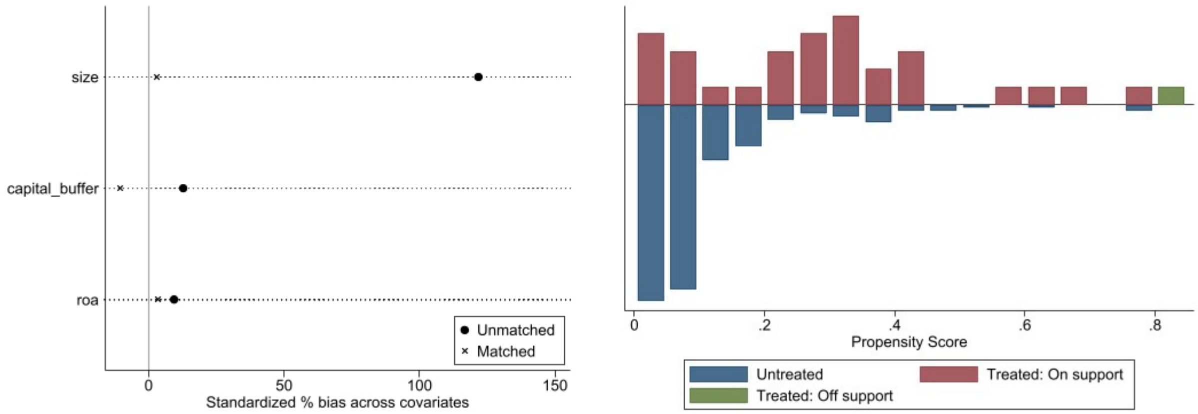

Figure 15: Propensity score matching efficiency

Table 11: **The effects of the mandatory retention rules on RWA ratio (PSM DID)**

	Dependent variable: RWA ratio		
	(1)	(2)	(3)
	OLS	OLS	PSM
Treatment	-0.011	0.004	0.021*
	(0.015)	(0.006)	(0.011)
Bank controls	No	Yes	
Number of bank	229	228	140

Notes: This table reports the effects of mandatory retention rules on banks' RWA ratios and delinquency rates under propensity score matching. There are missing values in the core deposits ratios, and they are dropped rather that imputed in the matching. Columns 1 and 2 are the benchmark cases without matching. Columns 3 and 4 report the PSMDID estimates using different sets of covariates. *Treatment* equals 1 if bank i is an RMBS-only securitizer after December 2015 and 0 otherwise. Standard errors are in parentheses. The superscripts ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.