Mandatory Retention Rules and Bank Risk *

Yuteng Cheng†

Bank of Canada

February 29, 2024

Abstract

This paper provides a theoretical model to address two questions: 1) how does mandatory retention affects banks' behavior in securitization, and 2) whether the current regulatory level of retention is optimal. In the model, retention strengthens borrower screening, but beyond a certain threshold, banks may start to shift risk, leading to unintended consequences. In addition, an empirical analysis of Dodd-Frank's retention rules suggests that RMBS banks both increased screening and shifted risk. The findings highlight that the current US retention rate of 5% is too high, as as the damage from risk shifting could outweigh the benefits of induced screening.

Key words: Securitization, Mandatory Retention, Risk Shifting

JEL codes: G21 G28

Declaration of interests: none.

* This paper is derived from a chapter of my PhD dissertation completed at the University of Wisconsin-Madison. 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, Stephen Miller, Narayan Bulusu, Jason Allen, Jonathan Chiu, Anson Zhou, David Lynch 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.

†Email: ycheng@bank-banque-canada.ca. Address: 234 Wellington St., Ottawa, ON, Canada.

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 screen 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 the role of banks 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. The main contribution is developing a theoretical model to investigate how mandatory retention affects banks' behaviors in securitization in general, and the corresponding optimal retention ratio. Additionally, using model predictions, I conduct an empirical exercise to explore whether the current regulatory level of retention is optimal.

To this end, I study two types of models: the main one with ex ante screening, more reflective of home loan securitizations, and another with ex post monitoring, occurring after the securitization process. The main model with bank screening demonstrates that retention strengthens screening as intended, but beyond a certain threshold, banks start to shift risk, leading to unintended consequences. There is a unique relationship between retention and bank behavior: as retention ratio rises, banks initially intensify their screening efforts, followed by a subsequent shift towards risk-taking activities. My empirical analysis of Dodd-Frank's stricter retention rules shows that banks that securitize mortgage loans

[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 (Flynn, Ghent, and Tchistyi (2020), Chemla and Hennessy (2014)), 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%.

both increased screening and shifted risk after the rules were implemented, suggesting that the current US retention rate of 5% is too high, as it may exacerbate risk-shifting behavior among banks. In the model that incorporates ex post monitoring, however, the effects of retention rules are ambiguous. A higher retention ratio does not uniformly lead to increased risk shifting. This variation adds complexity to the interpretation and implications of related policies for other types of securitizations.

In the main model, upon choosing a portfolio to fund, the bank securitizes and sells to investors a fraction of their portfolio, subject to a retention requirement. The bank's credit riskiness is determined by two types of moral hazard. One 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 portfolio with a low-NPV "gambling" portfolio that yields higher private returns if it succeeds. The bad portfolio thus contains higher credit risk. While the downside risk is absorbed by debtholders, the private returns go to shareholders if the investment strategy pays off. Hence, risk is shifted toward debtholders. The other is *costly screening effort* before the portfolio choice stage, which determines the quality of the pool of portfolios. The bank cannot commit to choosing the good portfolio as well as the screening effort.

In equilibrium, the retention constraint is binding. The retention ratio hence interacts with the two-dimensional moral hazard in the following way: As the ratio rises, the bank's value drops, since the financial benefits from securitization decrease, and this impact is more pronounced if the bank chooses the good portfolio. Meanwhile, screening efforts lead to an increase in the upside return of the portfolio, and the expected value of the gambling portfolio increases more than that of the good portfolio. In this context, screening and gambling are mutually reinforcing.

As a result, at a low retention ratio, the bank chooses to invest in the good portfolio because it is easier to securitize and market rates are favorable, but has little incentive to screen. As the retention ratio increases, banks begin to screen, and due to the bank's limited liability and the complementarity between screening and risk shifing, there is a tendency to shift towards riskier portfolios once the retention ratio surpasses a certain threshold.

The socially optimal retention ratio is strictly positive. Mandatory retention has three effects: it induces screening, it may also encourage risk shifting, and it reduces the gain from trade in securitization. Welfare is maximized when good portfolios are selected and screening 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 screening compared to a scenario without mandatory retention. Conversely, if we observe a simultaneous increase in screening and risk shifting, it indicates that the current retention ratio is too high.

I also extend the model to discuss the implications of my results for the form of optimal retention. In the retention requirements, banks can retain either a horizontal interest, consisting of the most subordinated tranches, or a vertical interest in each class of ABS tranches. I show first that the trade-off between screening 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.

In contrast, the model involving ex post monitoring leads to different policy implications. The primary difference stems from the fact that the retention constraint isn't always binding. There will be a jump in the payments from securitization investors as they anticipate monitoring. Hence, banks may want to retain more to force themselves to monitor even when the mandatory ratio is relatively low. Given this potential for a non-binding retention constraint, the complementarity between risk shifting and monitoring disappears. Consequently, a higher retention ratio does not necessarily result in increased risk shifting. In this scenario, while the optimal retention ratio remains an interior value, determining whether the retention is excessively high or low becomes challenging if increased risk shifting is observed.

Lastly, in an empirical analysis, I employ the Dodd-Frank Act as a quasi-experiment and use a difference-in-difference estimations strategy to examine the behavior of US bank holding companies (BHC) who are securitizers. The final rule of mandatory retention in Dodd-Frank was implemented at different times for different types of securitization. Specifically, for residential mortgage-backed securities (RMBS) — the focal point of the Dodd-Frank Act and a sector primarily concerned with borrower screening — these rules came into effect in

December 2015. For other securitization categories, the implementation occurred one year later. The treatment group hence contains BHCs that are RMBS securitizers, while the control group contains those that are not.

To assess risk shifting and screening, I utilize two distinct risk measures. The first is the ratio of risk-weighted assets (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. 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 borrowers after the implementation of the retention rules. Overall, according to the model prediction, these patterns happen only when the current rate of 5% is overshooting.

Ultimately, the findings of this paper not only guide regulators and practitioners in the banking and financial sectors towards more balanced securitization policies but also pave the way for future research to further refine these strategies and explore their long-term impacts on financial stability.

Related literature. This paper connects several different strands of literature. The first literature studies how securitization negatively affects banks' traditional roles. From a theoretical perspective, 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, works like Keys et al. (2010) and Purnanandam (2011) show that securitization led to lax screening standards of mortgages, while Piskorski, Seru, and Vig (2010) and Agarwal et al. (2011) provide evidence that securitization damaged servicing of loans, in particular renegotiation of delinquent loans. This paper contributes to this body of literature by linking these traditional roles to moral hazard in investment choices. Additionally, it distinguishes between screening and monitoring, demonstrating that mandatory retention has differing implications depending on whether the focus is on screening or monitoring.

The second literature discusses the structure of optimal retention, specifically examining which tranches banks should retain under retention requirements. Works by Fender and Mitchell (2009) and Kiff and Kisser (2014) discuss the effectiveness of retaining equity and mezzanine tranches in maximizing banks' screening efforts. Pagès (2013) argues that for optimal delegated monitoring by banks, a cash reserve account is more effective than retaining residual interest in securitization schemes. Malekan and Dionne (2014) study the optimal contract with regard to retention in the presence of moral hazard in a model with lender screening. These studies typically assume a fixed retention requirement. In contrast, my research investigates the determination of the optimal requirement, with findings applicable across various forms of retention. Additionally, the study of the optimal retention form is contextualized within the broader framework of moral hazard, offering a comprehensive perspective on this critical regulatory issue.

This paper also contributes to the extensive literature on risk shifting, a concept first articulated by Jensen and Meckling (1976). For a comprehensive early review of this literature, see Gorton and Winton (2003). Research by Keeley (1990), Demsetz, Saidenberg, and Strahan (1996), and Repullo (2004) investigates how different model setups for capital requirements can lead to diverse risk-taking behaviors among banks. These studies illustrate that the relationship between capital requirements and risk-taking is multifaceted, sometimes leading to reduced risk or, conversely, greater risk-taking. Notably, the work of Besanko and Kanatas (1996) is particularly relevant. Their examination of the interaction between managerial behavior, risk shifting, and capital requirements uncovers that increased capital requirements do not always reduce risk and may even enhance it. The two types of moral hazard settings they explore closely mirror the environment in my study. Building on this foundation, my paper aligns with and extends this body of literature by examining how another form of regulation—mandatory retention requirements in securitization—affects risk-taking behavior. While the existing literature primarily focuses on capital requirements, this study sheds light on the nuanced impacts of retention requirements on banks' risk-shifting behavior, thereby contributing to a broader understanding of regulatory impacts on banking risks.

Furthermore, this paper adds to the relatively limited empirical literature examining the impacts of retention rules. Furfine (2020) shows that after the implementation of the Dodd-Frank retention rules, loans in the CMBS market became safer, as indicated by metrics like interest rates, loan-to-value ratios, and income-to-debt-service ratios. Similarly, Agarwal et al. (2019) found that underwriting standards in the CMBS market tightened after the enforcement of these rules. My research complements these findings by also highlighting the phenomenon of risk-shifting, offering a more comprehensive view of the consequences of retention rules. [2]

Layout. The paper proceeds as follows. Section 2 describes the institutional background of securitization and risk retention. Section 3 introduces the main model with screening. Section 4 provides the equilibrium analysis. Section 5 extends the model and presents an alternative model that involves monitoring. Section 6 tests model predictions. Section 7 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. 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.

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

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

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.

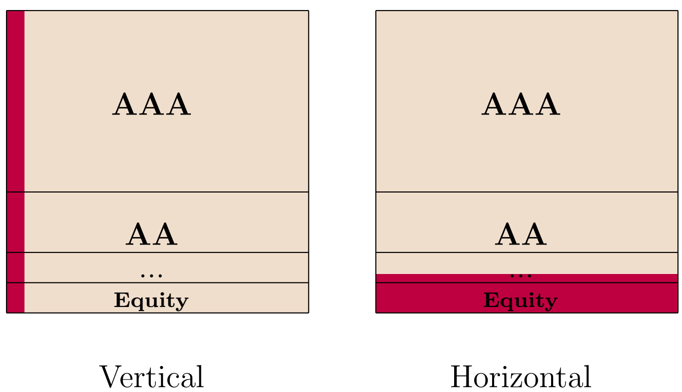

Figure 1: Retention options

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 Model with ex ante screening

This section presents the description of 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.

Consider an economy with a bank and multiple investors, both assumed to be risk-neutral. The economy lasts for four dates: $t \in \{-1, 0, 1, 2\}$. The bank needs to invest one unit of money in one of the two portfolios: a good portfolio and a bad (gambling) portfolio, the details of which will be defined later, using equity e and debt $d = 1 - e$, with the debt carrying an interest rate set at zero. At $t = -1$, the bank exerts effort $S \in \{0, 1\}$ in screening the two portfolios. Positive effort costs c, and the choice is unobservable by the market. At $t = 0$, the bank then chooses one of two portfolios: a good portfolio g and a gambling (bad) portfolio b. At time 1, the bank securitizes a fraction of the portfolio to investors. The securitization process is regulated by mandatory retention requirements, which will also be specified below. At $t = 2$, returns are realized. The timeline of the model is in Figure 6.

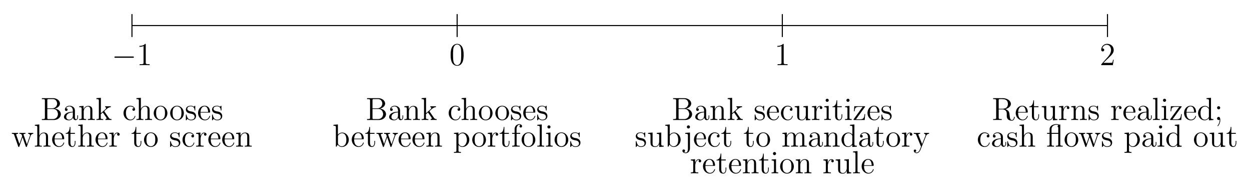

Figure 2: Timeline of the events

At time 0, the bank chooses which loan portfolio to invest in: $K \in \{g, b\}$. The expected return of a portfolio depends on screening, but for any level of screening, the good portfolio g has higher expected returns. Meanwhile, the bad portfolio b yields higher private returns if the gamble pays off. More specifically, given a screening level $S \in \{0, 1\}$, the portfolio payoffs are

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

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

$$q_g(S) = p_g - (1 - S)\Delta,$$

$$q_b(S) = p_b - (1 - S)\Delta,$$

with $p_g > p_b$. To clarify, when the bank opts to screen ($S = 1$), the success probabilities for the portfolios are just p_g and p_b, and the portfolio payoffs are

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

When the bank chooses not to screen ($S = 0$), the probability of success for a portfolio drops by Δ, and the portfolio payoffs are

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

The bad portfolio's success probability is always lower, $q_b(S) < q_g(S)$ for any S.[3]

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 $p_g = 1$.[4] The stochastic returns of the two portfolios are then assumed to satisfy

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

[3]The model's binary structure for screening effort can be extended to a continuous range. Specifically, the success probability could be represented as Sp_K, where $S \in [0, 1/p_g]$, as in Besanko and Kanatas (1996). Adopting this continuous framework would not change our results qualitatively.

Furthermore, the screening technology can be adapted to vary across different portfolios. For example, a good portfolio may require less screening compared to a bad one, in the sense that when $S = 0$, the drop of success probability Δ satisfies $\Delta_g < \Delta_b$.

[4]This assumption implies that when there is monitoring, the good portfolio is riskless.

The first inequality corresponds to the fact that the gambling portfolio yields a higher payoff in the success state, and the second inequality corresponds to the fact that the good portfolio has a higher expected return at any monitoring level.[5] The last inequality states that both portfolios have positive NPV. If the bank is solvent at time 2, it repays d to debtholders.

At time 1, the bank securitizes an exogenous fraction α of its portfolio.[6] The securitized part, refer to as the 'securitization pool', generates cash flows αR_K with probability $q_K(M)$, and is regulated by *mandatory retention requirements*: the bank must retain *no less than* θ fraction of the market value of the securitization pool, θ being set by the policymaker. In the vertical retention we focus on in this section, the bank holds a fraction Z of the pool; hence, the retention constraint is simply

$$Z \geq \theta.$$

Competitive investors observe the bank's portfolio choice K and retention choice Z, and bid a price (schedule) P.[7] However, the screening effort is unobservable. But investors are assumed to be sophisticated enough to take into account the bank's incentive problem. The total monetary benefit of securitization for the bank at time 1 is λP, where $\lambda > 1$.[8] 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).[9]

---

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

[6]I assume the bank does not securitize everything, so that the residual unsecuritized portfolio's payoff satisfies $(1 - \alpha)R > d$. That is, should the bank succeeds, it possesses sufficient funds to fulfill its debt obligations.

[7]The portfolio 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.

[8]λ is the gain from trade. This gain can be attributed to various factors such as reduced funding costs for new loans, access to unique investment opportunities, or an increase in reported profits linked to executive compensation. For example, the bank gains exclusive access to a new investment opportunity that yields private benefits $\lambda > 1$ per unit of investment. These elements, while not modeled in detail, underscore the multifaceted benefits securitization offers to the bank. Moreover, $\lambda > 1$ is equivalent to the bank discounting future cash flows at a higher rate than investors (e.g., DeMarzo and Duffie, 1999).

[9]Assume the bank repays an amount x of debt. The situation becomes equivalent to the bank receiving $\lambda(P - x)$ from securitization while being required to repay $d - x$ at time 2.

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

Definition 3.1 (Equilibrium.) *The equilibrium consists of the bank's screening choice S at time -1; the bank's portfolio choice K at time 0; the bank's retention choice Z at time 1; price schedule $P(K, Z)$ of the tranches sold to investors; such that given $P(K, Z)$, the bank chooses S, K, Z optimally subject to the retention constraint at time 1, and given S, $P(K, Z)$ is equal to the true value of the securities.*[10]

4 Equilibrium Bank Behavior

4.1 Securitization problem at $t = 1$

In vertical retention, investors purchase from the bank $1 - Z$ fraction of the pool and share the default risk of the underlying portfolio evenly in terms of seniority with the bank. If the bank goes bankrupt, the debtholders will take over the assets. For a given portfolio K, the schedule of prices investors offer based on Z is

$$P(K, Z) = (1 - Z)\alpha(p_K - \Delta^I)R_K$$

where $p_K - \Delta^I$ is what investors think will be the success probability since S is not observable, which will not be affected by the bank's choice of S.

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

$$\Pi_1(S, K; \theta) = \max_{Z} \ \lambda P(K, Z) + q_K(S)\Big((1 - \alpha)R_K + Z\alpha R_K - d\Big) - cS$$
$$\text{s.t. } Z \geq \theta.$$

[10]Under the provided definition of equilibrium, the model does not adopt a signaling game framework where the sender, instead of nature, selects the unobservable characteristics alongside other observable actions. This approach to signaling games often leads to an abundance of equilibria, complicating the analysis. For a detailed discussion on this issue, see In and Wright (2018). Moreover, Chemla and Hennessy (2014) explores a scenario where screening is combined with a follow-up possible signaling process through securitization structures.

Here, the total monetary benefits of securitization are represented by the product of λ and $P(K, Z)$. 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. The next result characterizes the bank's retention choice.

Proposition 4.1 (Retention choice under vertical component.) *The bank chooses at* $t = 1$

$$Z(S, K; \theta) = \begin{cases} \theta & \text{if } \frac{q_K(S)}{p_K - \Delta^I} \leq \lambda, \\ 1 & \text{if } \frac{q_K(S)}{p_K - \Delta^I} > \lambda. \end{cases}$$

In equilibrium, $p_K - \Delta^I = q_K(S)$, which implies the retention constraint will be binding: $Z = \theta$. When there is gain from trade as $\lambda > 1$, the bank simply retains the minimum possible amount.

With this result, the bank's value function at $t = 1$ can be written as

$$\Pi_1(S, K; \theta) = \lambda(1 - Z(S, K; \theta))\alpha(p_K - \Delta^I)R_K + q_K(S)\Big((1-\alpha)R_K + Z(S, K; \theta)\alpha R_K - d\Big) - cS.$$

4.2 Risk-shifting problem at $t = 0$

In this section, I study the bank's risk-shifting motives in the presence of limited liability, which imposes more costs on creditors in bad outcomes. I show that for any given screening level S, mandatory retention increases the bank's risk-shifting propensity, defined by the difference between the values of gambling and good portfolios

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

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

Proposition 4.2 (More retention leads to more risk shifting for given level of screening.) *In equilibrium, $G(\theta, S)$ is increasing in θ for any S.*

At a low retention ratio, the bank chooses the good portfolio because it has favorable market value and hence is easy to securitize; as the retention ratio increases, the value of choosing the good portfolio is hurt more because of the higher gain from trade generated, and this is the driving force of $G(\theta)$ being monotonically increasing in θ.

The next result presents a crowding-out theory of screening and risk shifting. I show that screening and risk shifting can be mutually reinforcing in equilibrium.

Proposition 4.3 (More screening leads to more bank risk shifting.) *In equilibrium,* $G(\theta,1) > G(\theta,0)$.

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

4.3 Screening at $t = -1$ and equilibrium characterization

In this section, 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 $G(\theta,0)$ in equilibrium. I also define $\bar{\theta}$ as the threshold satisfying $G(\bar{\theta},1) = 0$, or

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

Above this threshold, the bank always chooses to gamble if it has screened the pool of portfolios.

At the boundary $\theta = 1$, there are two possibilities for $G(1,0)$ (which denotes for $G(\theta = 1, S = 0)$). I first discuss the scenario when $G\big(1,0\big) \leq 0$, which means that gambling will not happen without screening. As mentioned earlier, a lack of due diligence in screening causes the value of the gambling portfolio 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 portfolio, due to the fact that

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

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

$$(1-\alpha)R_g + \alpha Z(1,g,\theta)R_g - d - c \geq (1-\Delta)\Big((1-\alpha)R_g + \alpha Z(0,g,\theta)R_g - d\Big) \qquad (1)$$

The bank is comparing profits conditional on success. The left-hand side is the expected profits under screening. Note that the payment from securitization purchasers is independent of the bank's screening effort. Therefore, it does not show up in the comparison. The right-hand side is the profits without screening, in which case the probability of success drops to $1-\Delta$. In equilibrium, $Z = \theta$, the above condition (1) simplifies as

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

This is the intended consequence of mandatory retention: the bank increases its effort in screening as the retention ratio rises.

Figure 3 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 $G(\theta,1)$.

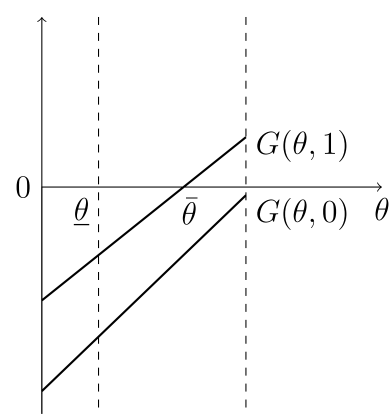

Figure 3: Risk-shifting propensity under screening choice

For θ greater than $\bar{\theta}$ but smaller than 1, the bank chooses the gambling portfolio if there is screening and chooses the good portfolio 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. The next result provides a complete characterization of the equilibrium.

Proposition 4.4 ***(Equilibrium.)*** *Assume $\bar{\theta} > \underline{\theta}$. When $\theta < \underline{\theta}$, the bank chooses the good portfolio $K^* = g$ but does not screen ex ante $S^* = 0$. When $\underline{\theta} \leq \theta \leq \bar{\theta}$, the bank chooses the good portfolio and screens $K^* = g, S^* = 1$. When $\theta \in (\bar{\theta}, 1]$, the bank screens $S^* = 1$ and gambles $K^* = b$. The retention constraint is always binding. Investors pay $P^* = (1-\theta)\alpha q_{K^*}(S^*)R_{K^*}$ accordingly.*

The assumption that $\bar{\theta}$ is greater than $\underline{\theta}$ generally holds true in scenarios where the screening cost, c, is relatively low. If the situation were reversed, $\underline{\theta} > \bar{\theta}$, the bank chooses the good project without screening up until $\underline{\theta}$. Beyond this point, the bank would engage in both screening and risk-taking behaviors. Under these conditions, the circumstance where the bank selects the good project and conducts screening concurrently does not arise. While this inverse scenario presents an intriguing discussion, my focus is on examining a model type where achieving the ideal scenario is feasible. Consequently, for the purpose of this study, I will set aside the condition where $\underline{\theta} > \bar{\theta}$.

In summary, within the entire policy parameter space $\theta \in [0,1]$: at $\theta = 0$, the bank chooses the good project without ex ante screening; as θ increases, the bank starts to screen at and beyond $\underline{\theta}$, and starts to shift risk at and beyond $\bar{\theta}$. This result is the combination of the dual insights derived from the preceding section: as the securitization regulation, as denoted by a higher θ value, the value of choosing the good portfolio is hurt more because of the higher gain from trade generated; meanwhile, risk shifting and screening are complements.[11]

The remaining case is when $G(1,0) > 0$, which is depicted in Figure 4.

[11]When the screening effort is continuous and the success probability takes the product form of Sp_k, the condition $G > 0$ function is irrelavent to the screening choice S. One can then establish directly the result that higher retention ratio leads to more risk shifting.

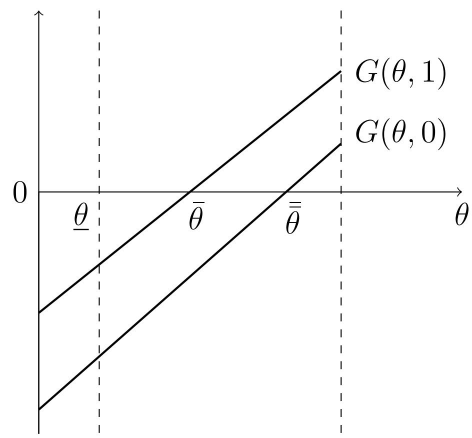

Figure 4: Risk-shifting propensity under screening choice

Let $\bar{\bar{\theta}}$ be such that $G(\bar{\bar{\theta}},0)=0$, which will be larger than $\bar{\theta}$. 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

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

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 outcome is the same.

4.4 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 portfolio being successful. In other words,

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

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

where

$$p(\theta) = \begin{cases} 1-\Delta & \text{if } \theta \le \underline{\theta}, \\ 1 & \text{if } [\underline{\theta}, \bar{\theta}], \\ p_b & \text{if } \theta > \bar{\theta}. \end{cases}$$

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

$$W(\theta) = \big(1 + (\lambda-1)\alpha(1-\theta)\big)p(\theta)R(\theta) - c(\theta),$$

which is the sum of total output and the net gain from securitization less the potential monitoring cost. The shape is shown in Figure 5.

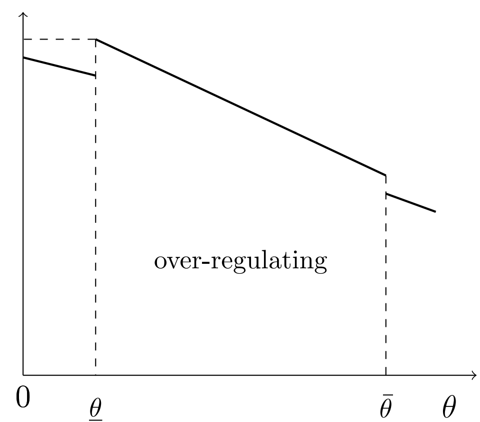

Figure 5: Welfare

There are jumps at $\underline{\theta}$ 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. The middle piece is the steepest, while the comparison of the slope of the other two pieces depend on which one is larger, $1-\Delta$ or p_b. In fact, if we impose the following assumption, the welfare is maximized at $\underline{\theta}$ and maximized when the good portfolio is chosen and screening effort is delivered:

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

This assumption is only about the good portfolio; hence, it is easy to satisfy when, say, R_g

is large enough, or c is small. 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 portfolio.

Proposition 4.5 (Optimal retention.) *The optimal retention ratio is $\underline{\theta}$. At this level, the bank screens and invests in the good portfolio.*

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 $\underline{\theta}$ and $\bar{\theta}$? In this region, the bank still screens and chooses the good portfolio; however, its securitization activities are over-regulated. In summary, welfare is maximized when the good portfolio is selected and just monitored, without securitization activities being overly regulated.

One remark is worth making. 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 $[\underline{\theta}, \bar{\theta}]$. This interpretation is equivalent to the scenario when there is huge social cost associated with bankruptcy, ξ, if the bank fails.

4.5 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.

	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 Section 6, banks do monitor more and shift more risk after the retention rule's implementation, implying that the current rate of 5% is overshooting.[13]

5 Extensions and discussions

This section expands the main model to cover different retention options and introduces a setup where the bank chooses monitoring after securitization instead of screening before it.

5.1 Horizontal retention

The model above assumes that the bank retains the vertical interest. The next two sections study two questions: 1) what are the bank's behaviors under horizontal retention, 2) the security design problem: which retention form is optimal from the bank's perspective. I will be using the h superscript to denote values in the case of horizontal retention.

In a horizontal retention, the bank retains Z fraction of the pool as the most junior tranche. 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 S, which is $q(S)\alpha R + (1 - q(M))\alpha r$, can be rewritten as

$$\alpha R - \big(1 - q(S)\big)\frac{R - r}{R}\alpha R,$$

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

[13]Note that if Assumption 1 fails, the socially optimal retention ratio is 0. This does not change this implication.

risk of the underlying portfolio 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

$$Z \leq \frac{R-r}{R}. \tag{2}$$

Given S, in equilibrium, the price schedule offered by the investors is

$$P^h(K,Z) = \begin{cases} (1-Z)\alpha R & \text{if } Z \geq \frac{R-r}{R}, \\ (1-Z)\alpha R - (1-q(S))\alpha\big(\frac{R-r}{R} - Z\big)R & \text{if } Z \leq \frac{R-r}{R} \end{cases}$$

The bank's screening decision affects the probability of investors receiving losses, which is $1 - q_K(S)$. 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 $(1-Z)\alpha R$. The value of the retained tranches, denoted by $V^h(K,Z)$, is the total value of the pool minus the value of the tranches sold: $V^h(K,Z) = V^{\text{pool}} - P^h(K,Z)$. Hence,

$$V^h(K,Z) = \begin{cases} \alpha ZR - \alpha(1-q(S))(R-r) & \text{if } Z \geq \frac{R-r}{R}, \\ \alpha q(S)ZR & \text{if } Z \leq \frac{R-r}{R}. \end{cases}$$

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

$$\begin{cases} ZR \geq \theta[q(S)R + (1-q(S))r)] + (1-q(S))(R-r) & \text{if } Z \geq \frac{R-r}{R}, \\ Zq(S)R \geq \theta[q(S)R + (1-q(S))r] & \text{if } Z \leq \frac{R-r}{R}. \end{cases}$$

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 $r > 0$, in equilibrium, similar to the main model, the bank's objective function is decreasing in Z. Once again, the retention constraint will be binding:

Proposition 5.2 (Retention choice.) *In equilibrium, under horizontal retention, the bank's retention choice of Z given (S, K) is*

$$Z^h(\theta) = \begin{cases} \frac{\theta[q(S)R+(1-q(S))r]+(1-q(S))(R-r)}{R} & \text{if } \theta > \frac{q(S)(R-r)}{q(S)(R-r)+r}, \\ \frac{\theta[q(S)R+(1-q(S))r]}{q(S)R} & \text{if } \theta \leq \frac{q(S)(R-r)}{q(S)(R-r)+r}. \end{cases}$$

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

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 only difference between this value under horizontal retention and vertical retention is that the bank's value function $\Pi_1(S, K; \theta)$ is still decreasing in θ but will have a kink point at $\theta = \frac{q(S)(R-r)}{q(S)(R-r)+r}$. Let

$$\theta_k = \min\left\{\frac{q(0)(R_g - r)}{q(0)(R_g - r) + r}, \frac{q(0)(R_b - r)}{q(0)(R_b - r) + r}\right\}.$$

A similar analysis can be conducted for the portfolio choice and screening stage on $[0, \theta_k]$, obtaining the same mechanism that more retention leads to more screening as well as risk shifting.

Proposition 5.3 ***(Equilibrium.)*** *There exist two thresholds $\underline{\theta}^h$ and $\bar{\theta}^h$ such that the bank chooses the good project without screening if $\theta < \underline{\theta}^h$, the bank chooses the good project and screens if $\theta \in [\underline{\theta}^h, \bar{\theta}^h]$, and the bank gambles and screens if $\bar{\theta}^h < \theta \leq \theta_k$.*

In particular, upon choosing the good project, the bank is indifferent between screening and not screening when

$$(1-\alpha)R_g + \alpha\theta R_g - d - c = (1-\Delta)\left((1-\alpha)R_g + \alpha\frac{\theta((1-\Delta)R_g + \Delta r)}{(1-\Delta)} - d\right).$$

Therefore the new threshold that induces screening is

$$\underline{\theta}^h = \frac{\frac{c}{\Delta} + d - (1-\alpha)R_g}{\alpha(R_g - r)}.$$

Note that this threshold is higher than the one under vertical retention because as the bank retains more, its payoff without screening increases, making it harder to motivate screening.

The predictions in Section 4.5 continue to hold: if one observes a simultaneous increase in screening and risk shifting of banks in the data, it means the retention ratio is higher than $\bar{\theta}^h$, and lowering such ratio can increase welfare.

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 portfolio K, if capital requirements are not a concern, or "slack" for the bank, the expected value under horizontal retention is higher than under vertical retention. On the one hand, when the retention constraint is binding, the price paid by investors is the same under the two retention forms, which is equal to the $1-\theta$ fraction of the pool's market value

$$P = (1-\theta)\alpha(qR + (1-q)r).$$

On the other hand, 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.

Proposition 5.4 *If the capital requirement is slack at retention requirement θ, the bank chooses horizontal retention.*

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[14] 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 Alternative model with ex post monitoring

In this section, I present an alternative version of the model incorporating ex post monitoring. In this variation, although an optimal interior retention ratio persists, relationship between the retention ratio and the incentive for risk-shifting becomes non-monotonic, leading to complex policy implications.

The economy lasts for four dates: $t \in \{0, 1, 2, 3\}$. At $t = 0$, the bank faces the same portfolio choice problem. At time 1, the bank securitizes a fraction of its portfolio to investors. Now, at time 2, the bank chooses whether to monitor its portfolio $M \in \{0, 1\}$, where $M = 1$ denotes monitoring and $M = 0$ denotes no monitoring.[15] The monitoring effort affects the credit risk of the portfolio as a whole, not just the portion the bank retains. Specifically, similar to the screening case,

$$q_K(M) = p_K - (1 - M)\Delta.$$

At time 3, returns are realized. The timeline of the model is in Figure 6.

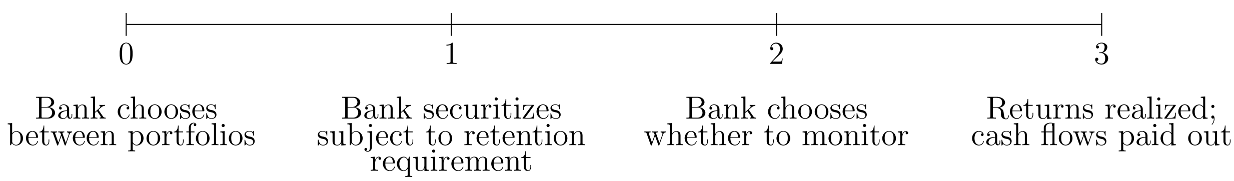

Figure 6: Timeline of the events

[14]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.

[15]In practice, monitoring includes collecting payments, renegotiating, and working closely with the trustee representing investors' interests.

Same as in the main model, I assume $r = 0$ and vertical retention. I introduce an additional assumption for the model.

Assumption 2 $c > \lambda\big(\Delta R_b - p_b d\big)$.

This assumption states that the monitoring cost is not too small for the bad project, making monitoring indeed a friction. As shown later, if this assumption is violated, banks will voluntarily choose to monitor, even when there is no retention policy. Since $R_b > R_g$ and $p_b < p_g$, this assumption also holds for the good project.

I present the definition of equilibrium in this model.

Definition 5.1 (Equilibrium.) *The equilibrium consists of the bank's portfolio 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 make investors break even and the bank maximizes its value sequentially subject to retention constraint at time 1.*

The model is solved backwards. I first solve for the bank's monitoring decision at time 2, after the securitization stage. Given the portfolio 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 portfolio, $Z\alpha R$ 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 5.5 *Given (K, Z), the bank chooses $M = 1$ if and only if*

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

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." 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 portfolio return R_K. If the bank chooses the bad portfolio, it has more incentive to monitor.

In vertical retention, investors purchase from the bank $1 - Z$ fraction of the pool and share the default risk of the underlying portfolio evenly in terms of seniority with the bank. 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 portfolio, 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 5.6 *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 have enough incentive to monitor, the bank must retain at least $\bar{Z}_K$ of the portfolio. While monitoring improves portfolio quality, it comes at a cost: lowered securitization revenue due to retention and the cost of monitoring. 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 7 illustrates the bank's objective function under this assumption, with the second piece of the function being steeper than the first.

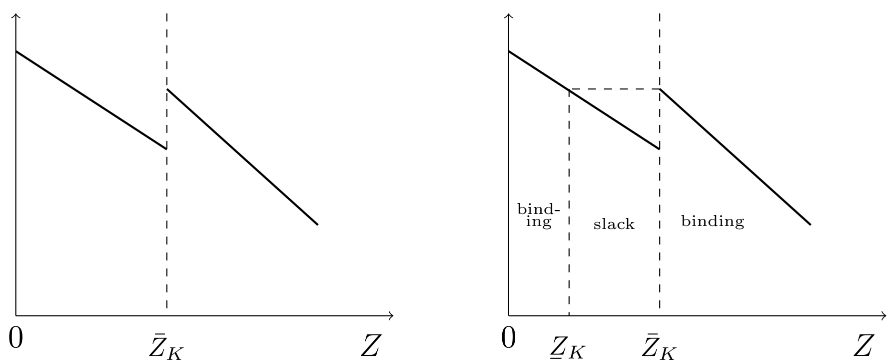

Figure 7: The bank's objective function and retention constraint

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 7. In fact, the formula of $\underline{Z}_K$ for portfolio K is

$$\underline{Z}_K = 1 - \frac{(\lambda p - (p - \Delta))(R - d) - (\lambda - 1)\frac{c}{\Delta} - c}{(\lambda - 1)\alpha(p - \Delta)R}. \tag{3}$$

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 observations are summarized by the following result.

Proposition 5.7 (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 5.8 (More retention leads to more monitoring.) *Given a portfolio K, the policymaker can set $\theta \geq \underline{Z}_K$ to implement monitoring. In particular, if the policymaker sets*

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

the bank always monitors.

For a given portfolio K, under the optimal retention choice, the bank's expected value $\Pi_1(\theta|K)$ is graphed in Figure 8. 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 portfolio has a larger $\underline{Z}$.

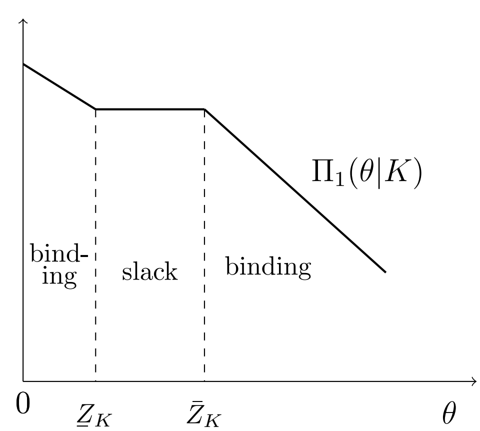

Figure 8: The bank's expected value under the optimal retention choice

Lastly, I complete the equilibrium characterization by studying the bank's risk-shifting motives. 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 portfolio choice.

At time 0, for a required retention ratio θ, let the bank's expected value of choosing the good portfolio 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 5.7. 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

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

Define again 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 risk-shifting gain function under the different requirements of θ. Define a threshold

$$\bar{\lambda} = \frac{\frac{c}{\Delta}(1-p_b)}{\frac{c}{\Delta}(1-p_b) - [p_b(R_b - d) - (R_g - d)]},$$

such that at $\lambda = \bar{\lambda}$, $V_g(\bar{Z}_g) = V_b(\bar{Z}_b)$. The main result of this alternative model is the following.

Proposition 5.9 (Equilibrium)

1. *If $\lambda > \bar{\lambda}$, there exist two thresholds $\underline{\theta}, \bar{\theta}$ with $\underline{\theta} < \bar{\theta}$ such that the bank chooses the good portfolio but does not monitor for $\theta < \underline{\theta}$; the bank chooses the good portfolio and makes an effort to monitor for $\theta \in [\underline{\theta}, \bar{\theta}]$; the bank gambles and monitors subsequently for $\theta > \bar{\theta}$.*

2. *If $\lambda \leq \bar{\lambda}$, there exist three thresholds $\underline{\theta}, \bar{\theta}, \bar{\bar{\theta}}$ with $\underline{\theta} < \bar{\theta} < \bar{\bar{\theta}}$ such that the bank chooses the good project without monitoring if $\theta < \underline{\theta}$; gambles and monitors if $\theta \in [\underline{\theta}, \bar{\theta})$; chooses the good project and monitors if $\theta \in [\bar{\theta}, \bar{\bar{\theta}}]$; gambles and monitors again if $\theta > \bar{\bar{\theta}}$.*

The results illustrate that the impact of retention on banks' behavior varies with the extent of securitization gains λ. In scenarios where the gain is substantial, the model aligns closely with the screening model. Conversely, when the gain is small, the monotonic relationship between retention and risk shifting disappears. This is due to the flat part of the bank's value function where the retention constraint is slack.

Consequently, the policy implications are not straightforward for small values of λ. Indeed, the optiaml level of retention is interior. However, when both monitoring and risk shifting are observed, it becomes challenging to ascertain whether the retention ratio is excessively high or low, as the bank can engage in gambles and monitors when θ is either below or above $\bar{\theta}$.

6 Empirical Evidence

In the last section, I examine the behavior of US securitizers after the Dodd-Frank retention rules were implemented. The main focus is the risk-shifting part, which is the side effect of the rentetion rules shown in the model. Overall, I show that there was a simultaneous increase in screening/monitoring and risk shifting of banks who specialize in residential mortgage securitization. Under the screening model framework, this happens only when the current rate of 5% exceeds the optimal level.

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. [16] 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

[16] 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 securitized with servicing retained or with recourse or other seller-provided credit enhancements in millions of US dollars.

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 portfolios, or engaging in risk shifting.[17]

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 difference-in-difference (DID) estimation 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. I compare 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.[18]

As mentioned in footnote 17, 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.

[17] 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.

[18] 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.

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.[19] 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:

$$RWA_{it} = \beta D_{it} + X_{it}'\delta + \alpha_i + \gamma_t + \epsilon_{it}, \tag{4}$$

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. Standard errors are clustered at the bank level. The coefficient of interest is β. I estimate (4) with and without bank-specific controls. It's important to note that banks have different shares of RMBS securitization activities, resulting in varying levels of treatment intensity. Hence, I also consider a continuous treatment study in Appendix C.

To shed light on the dynamics of the average treatment effect before and after the passage

[19]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.

of the mandatory retention rule, I replace the dummy D_{it} in specification (4) with leads and lags, which I name as $Before$ and $After$. The new regression model is

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

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 β 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.

The summary statistics for the treatment and control group are reported in Table 4 in Appendix B. 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 (0.78 versus 0.75). Though operating at a higher asset level, banks in the control group have more profitable portfolios than banks in the treatment group (0.009 versus 0.006). As a sign of bank solvency and stability, the treatment group banks are more capitalized than control group banks (0.14 versus 0.10).

Table 2: **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

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

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 2 reports the estimation of equation (4). The first two columns report the results when no other control variables are included. The treatment groups are defined as RMBS activity share ≥ 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 5 in Appendix B 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 3: **The dynamics of treatment effects**

	Dependent variable: RWA ratio		
	RMBS ≥ 0.5	RMBS ≥ 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

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

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 3, I report the estimation of equation (5) 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_M: the results in the previous table occur after the Final Rule's implementation. Figure 9 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 9: Time passage relative to the mandatory retention rule's implementation

Lastly, I also study the impact of the retention rules on bank delinquency rates. Results are presented in Tables 6 and 7 in Appendix B. Across all specifications, the results show that the delinquency rate decreased by approximately 0.3 percentage points.

In summary, loans are riskier ex ante and safer ex post. Consequently, banks shifted toward riskier investing strategies and must have engaged in more efforts to screen and monitor their borrowers. According to the screening model prediction, the current 5% ratio is overshooting.

7 Concluding Remarks

Policymakers have implemented new risk retention rules with the expectation of aligning incentives and reducing banks' risk. However, as this paper has demonstrated through both theoretical models and empirical analysis, the impact of these retention policies is complex and multifaceted. The main model of this study, with its focus on ex ante screening, indicates that while retention does enhance screening up to a point, it also leads to increased risk shifting beyond a certain threshold. Furthermore, my empirical findings reveal that the introduction of stricter retention rules led to an increase in both screening and risk shifting among banks securitizing mortgage loans. This suggests that the current retention rate of 5% in the U.S. may actually be too high, inadvertently exacerbating riskier behaviors among banks. These results highlight the delicate balance policymakers must strike in setting retention levels that effectively align banks' incentives with broader financial stability goals, without inadvertently promoting riskier behaviors.

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.

Besanko, David and George Kanatas. 1996. "The Regulation of Bank Capital: Do Capital Standards Promote Bank Safety?" *Journal of Financial Intermediation* 5 (2):160–183.

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.

In, Younghwan and Julian Wright. 2018. "Signaling Private Choices." *The Review of Economic Studies* 85 (1 (302)):558–580.

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.

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.

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

Proof of Proposition 4.1. The partial derivative of the bank's objective with respect to Z is

$$\alpha R(-\lambda(p_K - \Delta^I) + q_K(S)).$$

The result follows. □

Proof of Proposition 4.2. In equilibrium, $Z(\theta) = \theta$, and $p_k - \Delta^I = q_k(S)$. We have

$$\begin{aligned}\frac{\partial G}{\partial \theta} &= (\lambda - 1)\alpha\Big(q_g(S)R_g - q_b(S)R_b\Big)\\ &> 0,\end{aligned}$$

as $q_g(S)R_g > q_b(S)R_b$ for any S. □

Proof of Proposition 4.3. In equilibrium, $Z(S, K, \theta) = \theta$, we have

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

which completes the proof. □

Proof of Proposition 4.4. Since $\theta > \underline{\theta}$, screening generates a higher value for the bank undertaking the good portfolio. And under screening, the gambling portfolio dominates the good portfolio because $\theta > \bar{\theta}$. □

Proof of Proposition 5.1 If $r = 0$, $\bar{R} = pR$ and $\underline{R} = (1-\Delta)R$. The rest follows. □

Proof of Proposition 5.2. If $\theta > \frac{q(S)(R-r)}{q(S)(R-r)+r}$, then $\frac{\theta[q(S)R+(1-q(S))r]}{q(S)R} > \frac{R-r}{R}$, the bank cannot choose $Z \leq \frac{R-r}{R}$, it will choose the lowest possible Z, which is $Z = \frac{\theta[q(S)R+(1-q(S))r]+(1-q(S))(R-r)}{R}$; if $\theta \leq \frac{q(S)(R-r)}{q(S)(R-r)+r}$, the bank is able to choose $Z \leq \frac{R-r}{R}$, hence it chooses $Z = \frac{\theta[q(S)R+(1-q(S))r]}{q(S)R}$. □

Proof of Proposition 5.3. When $\theta \leq \theta^k$, the bank retains $Z(\theta) = \frac{\theta(q(S)R+(1-q(S))r)}{q(S)R}$ in equilibrium. It follows that

$$\begin{aligned}\frac{\partial G}{\partial \theta} &= \lambda\alpha(q_g(S)R_g - q_b(S)R_b) - \alpha(q_g(S)R_g + (1-q_g(S))r - q_b(S)R_g - (1-q_b(S))r)\\ &> (\lambda - 1)\alpha\Big(q_g(S)R_g - q_b(S)R_b\Big)\\ &> 0,\end{aligned}$$

where the second inequality comes from the fact that $q_g(S) > q_b(S)$ for any S. Meanwhile,

$$\begin{aligned}
G(\theta,1)-&G(\theta,0)\\
&= \lambda\alpha(1-\theta)p_bR_b + p_b\Big((1-\alpha)R_b + \frac{\theta(p_bR_b+(1-p_b)r)}{p_bR_b}\alpha R_b - d\Big)\\
&\quad - \lambda\alpha(1-\theta)R_g + \Big((1-\alpha)R_b + \theta\alpha R_g - d\Big)\\
&\quad - \lambda\alpha(1-\theta)(p_b-\Delta)R_b - (p_b-\Delta)\Big((1-\alpha)R_b + \frac{\theta((p_b-\Delta)R_b+(1-p_b+\Delta)r)}{(p_b-\Delta)R_b}R_b - d\Big)\\
&\quad + \lambda\alpha(1-\theta)(1-\Delta)R_g + (1-\Delta)\Big((1-\alpha)R_g + \frac{\theta((1-\Delta))R_g+\Delta r)}{(1-\Delta)R_g}R_g - d\Big)\\
&= \Delta(\lambda\alpha(1-\theta)+\theta+1-\alpha)(R_b-R_g)\\
&>0,
\end{aligned}$$

which remains unchanged. The rest of the argument follows the same steps in the main model. □

Proof of Proposition 5.5 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. □

Proof of Proposition 5.6 For $Z < \bar{Z}$, the bank does not monitor. The bank's objective function is

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

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

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

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. □

Proof of Proposition 5.8 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

$$p_b(R_b - d) > (R_g - d),$$
$$R_b > R_g.$$

The proof is complete. □

Proof of Proposition 5.9 First, we show that $G(\theta)$ is increasing in θ when the retention constraints are binding for both portfolios; 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,

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

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

Second, assume $\lambda>\bar{\lambda}$. 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,

$$\bar{\theta}=\frac{(\lambda\alpha+1-\alpha)(R_g-p_bR_b)-(1-p_b)d}{(\lambda-1)\alpha(R_g-p_bR_b)} \tag{6}$$

which is the same as in the main model. 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 **??**, 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 $(1-\Delta)R_g$, because the bank monitors the gambling portfolio but not the good one. We will assume later that $p_b>1-\Delta$, hence $p_bR_b>(1-\Delta)R_g$, and the maximum is achieved at $\bar{Z}_b$. At this point, by the definition of $\bar{\lambda}$, we have

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

Let $\underline{\theta}\equiv\underline{Z}_g$. Assumption **??** also guarantees that $\underline{\theta}<\bar{\theta}$. When $\theta\in[\underline{\theta},\bar{\theta}]$, the bank chooses the good portfolio and monitors.

Lastly, consider $\lambda\leq\bar{\lambda}$, then $V_b(\bar{Z}_b)>V_g(\bar{Z}_g)$. Meanwhile, by assumption **??**, $V_b(\bar{Z}_g)<V_g(\bar{Z}_g)$. Therefore, let $\underline{\theta}$ be such that

$$V_b(\underline{\theta})=V_g(\underline{Z}_g),$$

and let $\bar{\theta}$ be such that

$$V_b(\underline{Z}_b)=V_g(\bar{Z}_g),$$

and $\bar{\bar{\theta}}$ be such that

$$V_b(\bar{\bar{\theta}}) = V_g(\bar{\bar{\theta}}),$$

the results follow. $\square$

Appendix B Supplementary tables and graphs

Table 4: **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.

Table 5: **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 6: **The effect of the retention rule on banks' loan delinquency rate**

	(1) RMBS≥0.5	(2) RMBS=1	(3) 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 7: **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.

Appendix C Continuous treatment

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

$$RWA_{it} = \beta Share_{it} * D_t + X'_{it}\delta + \alpha_i + \gamma_t + \epsilon_{it}, \tag{7}$$

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

Table 8 displays the regression results of equation (7). 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.

Table 8: **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

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

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.